

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 16, 2003, 22:51:47 ; Search time 459 Seconds
(without alignments)
3158.166 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555
Perfect score: 2887
Sequence: 1 LPGSTPASVGRRLPKNPKTG.....HTQSSVPLPATSMSVSGMA 537

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DRV=xlp
-Q=/cgp2_1/USPTO.spool_p/US09712338/runat_16092003_144913_19119/app_query.fasta_1.711
-DB=N_Geneseq_19jun03 -QFMT=fastp -SUFFIX=p2n.rng -MATRIX=blom62 -TRANS=human40.cdi
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USER=US09712338_@CNG_1_1_490_@runat_16092003_144913_19119 -NCPH=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
- 25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	2883	99.9	1662	19	AAV28620	A. oryzae ATCC2038
2	2338	81.0	1656	24	ABZ78288	A. niger serine ca
3	1983	68.7	3150	24	ABZ78231	A. niger serine ca
4	980.5	34.0	1872	24	ABZ78289	A. niger serine ca
5	977	33.8	3221	24	ABZ78232	A. niger serine ca
6	941	32.6	1665	24	ABZ78283	A. niger serine ca
7	916.5	31.7	1581	24	ABZ78241	A. niger serine ca
8	894	31.0	2940	24	ABZ78184	A. niger serine ca
9	887.5	30.7	3080	24	ABZ78226	A. niger serine ca
10	704.5	24.4	1007	25	ABZ51997	A. niger serine ca
11	636	22.0	2027	24	ABZ76315	Aspergillus oryzae
12	630	21.8	1611	24	ABZ78243	S. cerevisiae BAX-
13	620	21.5	2002	17	AAT28284	A. niger carboxype
14	580	20.1	2068	17	AAT28283	A. niger SFAG 2 ca
15	580	20.1	2660	24	ABZ78186	A. niger Bo-1 geno
16	568.5	19.7	2503	15	AAO55347	A. niger carboxype
17	541.5	18.8	1653	24	ABO76547	Sequence of gene X
18	474	16.4	1551	24	ABZ12878	C. albicans BAX-as
19	460.5	16.0	678	21	AAF12116	Aspergillus thalia
20	414	14.3	1446	24	ABZ78269	Aspergillus oryzae
21	383	13.3	1510	21	AAC42522	A. niger carboxype
22	372	12.9	626	21	AAF12522	Arabidopsis thalia
23	365	12.6	1368	24	ABZ78267	Aspergillus oryzae
24	361.5	12.5	1936	21	AAC40187	A. niger carboxype
25	361	12.5	1551	19	AAV64076	Arabidopsis thalia
26	359	12.4	1686	22	AAF33808	Human serine carbo
27	359	12.4	1428	22	AAF94477	Human cDNA encodin
28	359	12.4	1633	22	AAF94487	Human hydrophobic
29	359	12.4	1695	21	AAA47444	Human hydrophobic
30	359	12.4	1963	22	AAH99751	Human RANGO 176 co
31	359	12.4	2076	22	AAH52243	Human protein enco
32	359	12.4	2076	22	AAH61129	Protein PRO223 CDN
33	359	12.4	2076	22	AAH72401	Human DNA encoding
34	359	12.4	2076	25	ACA54942	Human PRO223 cDNA
35	359	12.4	2076	25	ACA57887	Novel human secret
36	359	12.4	2076	25	ACA58427	Human PRO223 cDNA
37	359	12.4	2076	25	ACA60134	cDNA encoding huma
38	359	12.4	2076	25	ABX98357	Human cDNA for sec
39	359	12.4	2076	25	ABX98859	Human cDNA encodin
40	359	12.4	2076	25	ACA05472	Novel human secret
41	359	12.4	2076	25	ACA05904	cDNA encoding huma
42	359	12.4	2076	25	ABX96151	Human secreted/tra
43	359	12.4	2076	25	ABX97948	Human secreted/tra
44	359	12.4	2076	25	ABX78732	Human PRO polynucl
45	359	12.4	2076	25	ABX75745	Human PRO polynucl

ALIGNMENTS

RESULT 1
AAV28620
ID AAV28620 standard; DNA; 1662 BP.
AC AAV28620;
XX
XX
XX
DT 27-AUG-1998 (first entry)
XX
XX A. oryzae ATCC20386 carboxypeptidase I DNA.
DE
DE
KW Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous;
KW food industry; ss.
XX
XX Aspergillus oryzae.
XX
FH Key Location/Qualifiers
FT CDS 1..1662

```

FT      /*tag- a
FT      /product= carboxypeptidase I
FT      /note= "partial coding sequence"
XX      W09814599-A1.
XX      09-APR-1998.
XX      03-OCT-1997; 97WO-US17977.
XX      27-NOV-1996; 96US-0757534.
XX      04-OCT-1996; 96US-0726880.
XX      (NOVO ) NOVO NORDISK BIOTECH INC.
XX      (NOVO ) NOVO-NORDISK AS.
XX      Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;
XX      PI Klotz A, Mathisen TE, Rey M;
XX      WPI: 1998-240098/21.
XX      P-PSDB; AAW56099.
XX      Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous
XX      substrates, useful for improving flavour of foods
XX      Claim 2; Fig 3; 82pp; English.
XX      This DNA sequence encodes carboxypeptidase I from Aspergillus oryzae.
XX      This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
XX      25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a
XX      residual activity of at least 65.5% after 30 minutes at pH 4.0 and
XX      60 deg. C. It also has the capacity to hydrolyse x from N-CBZ-Ala-x where
XX      N-CBZ is N-carbobenzoxyl and X is any amino acid. The carboxypeptidases
XX      can be used for obtaining hydrolysates (which can be enriched in free
XX      glutamic acid or peptide bound glutamic acid residues) from proteinaceous
XX      substrates. The carboxypeptidases can be used in flavour-improving
XX      compositions in the food industry. The products can also be used for the
XX      production of polypeptides free of carboxypeptidase activity.
XX      SQ Sequence 1662 BP; 396 A; 468 C; 398 G; 400 T; 0 other;

Alignment scores:
Pred. No.: 3,5e-276 Length: 1662
Score: 2883.00 Matches: 536
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.86% Indels: 0
DB: 19 Gaps: 0

US-09-712-338-2_COPY_19_555 (1-537) x AAV28620 (1-1662)
QY      1 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 20
DB      55 CTTCAGAGAGTACACCGGGTCCGTCGGTAGAGACAGCTACCCAGAGACCCACCGGG 114
QY      21 VallyThrLeuThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu 40
DB      115 GTCAGACTCTTACACCGCAACAAATGTCACATCCGGTACAGGAACCCGGGGCAGAG 174
QY      41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60
DB      175 GCGCTCTGCGAGACTACCCGGGTGCAAAATCTCTCTGATATGTCGACACCTCTCCC 234
QY      61 GluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProfil 80
DB      235 GAGTCCCATACCTCTCTGTTCTCTGAGCCAGACATACCCAGAAATGACCTATC 294
QY      81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeu 100
DB      295 ACATGTGTGTAATGGTGGCCCTGGAAGCGATCTTTGATCGGTCTCTTCGAAGAGTTG 354
QY      101 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerIrpAsnGlu 120
DB      1435 TATGTAATTTCTCCTTCTCCTTCTATGAGCAGGCCATGAAGTCCCATACCTACCAG 1494

```

```

355 GGCCCTTGCCATGTCATCAATTCGAGCTTTTGTGACTACATCAACCCCTCAGTCGTGAACGAG 414
121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 140
415 GTCTCCAAATTACTATTCTCTGTCGCCAGCCATTGGGAGTCGGCTTTTCATATAGTATAG 474
141 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 160
475 GTTATGGGTCCATTAAACCTGTAACTGGGGTCTGCGAAATTCGAGCTTTGAGGAGTT 534
161 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu 180
535 CAGGGCGGTACCCCAACCAITGATGCCACTCTGATCGATACCTACCTAGCTTGGAGTTC 594
181 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 200
595 GCGCTTGGGAGATCTCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGAGTTC 654
201 GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 220
655 CAGCTTAAGGACTTCACTCTATGACGGAGAGTATGGAGGGCAGCTATGTCCTTCATTC 714
221 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 240
715 TTCAATCATTTTACGAGCAGAAAGAGAGAATTGCCAACGGTAGTGTATATGGTGTTCAG 774
241 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro 260
775 CTTAATTTCAACTCTCTGGGAATTATTACGGCATCATCGACGAGGCGATCCAGGCCCT 834
261 TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal 280
835 TACTACCTCAATTCGCTGTGACAAATACCTACCGTATCAAGGCTGTCAACGAGACCGTC 894
281 TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 300
895 TACAACATACATGAAGTTTGCCACCAATGCCAAATGGTTCAGGATTTGATTCCACC 954
301 CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 320
955 TGCAACACAGACAAACCCGCGATTAGCTGACTACGCGCTCTCGCGCAAGCCACCAAC 1014
321 MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 340
1015 ATGTCCAGGGACAAATGTGAGGGCCATCTACCGCTTGTCTGTCGGTGTGTATGAT 1074
341 IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 360
1075 ATTCGGCATCCATATGATGACCCGACTCCGCGCAAGTTATTACACAAATTTCTGCAAG 1134
361 AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 380
1135 GACTCTGTCTATGGAGCTATCGCGCTCAACATCAACTACACCTACCGCTCAATAATGACGTC 1194
381 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 400
1195 TACTACGCTTTCACGACAAACAGGCGACTTGTCTGGCCCAACTCATCGAAGACCTCGAG 1254
401 GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 420
1255 CAGATCTTGTCTCTCCCGTGGTGTCTCCCTCATCTATGCGAGCGCGGATTACATCTGC 1314
421 AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 440
1315 AACTGGTTCGGCGGTTCAGCGGCTTCCCTCGTGAACCTACTCCCAAGCCGCCAGTTTC 1374
441 ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 460
1375 CGAGCGCAGGGTACACGCCCTCTGAAGTCAACGGCTCGAGTATGGGGAACCTCGCAG 1434
461 TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 480
1435 TATGTAATTTCTCCTTCTCCTTCTATGAGCAGGCCATGAAGTCCCATACCTACCAG 1494

```


QY	326	-----	326	PN	WO200269623-A2.
DB	2005	CGGTGAGTGTCTACTGTTCTCTGCGGGGTGCAATGAUCAAGGACTTTGCTAAGCTG	2064	XX	06-SEP-2002.
QY	327	-----	343	XX	22-FEB-2002; 2002WO-EP01984.
DB	2065	TCATGTACAGAAGGCCCTACTACACAGTTTGGCGCGTGTATGATATCGGCAC	2124	XX	23-FEB-2001; 2001EP-0200657.
QY	344	ProTyrAsp	346	PR	23-FEB-2001; 2001EP-0200658.
DB	2125	CCCTACAA-TGTAACTGGCAAGATAAAGGATTTCTCCGAACAGGACACTGCTCAT	2183	PR	26-FEB-2001; 2001EP-0200706.
QY	347	-----	362	PR	26-FEB-2001; 2001EP-0200707.
DB	2184	ATGTCAACTAGAGCCCGCGCTCTACTTGTGACTACCTCAAGAAAGACTCA	2243	PR	26-FEB-2001; 2001EP-0200708.
QY	363	ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnValTyrTrp	382	PR	26-FEB-2001; 2001EP-0200719.
DB	2244	GTATGGATGCTATGCGCGTGGACATTAACACCGAGTCCAGCGCGGAAGTATATAT	2303	PR	28-MAR-2001; 2001EP-0000075.
QY	383	AlaPheGlnGlnThrGlyAspPheValTTPProAsnPheIleGluAspLeuGluIle	402	PR	28-MAR-2001; 2001EP-0000080.
DB	2304	GCATTCACGACAGCCGCGACTTTGTATGGCGAATTCATTGAGGACCTCGAAGAGATC	2363	PR	28-MAR-2001; 2001EP-0000087.
QY	403	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	422	PR	28-MAR-2001; 2001EP-0000098.
DB	2364	CTCCAACTCCCGTACCGGTGCTGTGATCTAGGCGGATGCCGACTATATCTCTAAGTGG	2423	PR	28-MAR-2001; 2001EP-0000156.
QY	423	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer	442	PR	21-MAY-2001; 2001EP-0000159.
DB	2424	TTGCGCGGTGAGGCCATCTACATCGCAGTTAACTAGCCCATGAGTCCGTCGCA	2483	PR	21-MAY-2001; 2001EP-0000160.
QY	443	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly	462	PR	21-MAY-2001; 2001EP-0000162.
DB	2484	GCGGGATACACCCCAATGAGTGGGTGCAATACCGTGGAGTCCGAGTCCGAGTAGGC	2543	PR	21-MAY-2001; 2001EP-0000165.
QY	463	AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle	482	PR	21-MAY-2001; 2001EP-0000166.
DB	2544	AACTTTCGTTCACCCGCGTATATCAGCGTGGCAGAGTTCCATATCTATCAACCGATC	2603	PR	21-MAY-2001; 2001EP-0000168.
QY	483	AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys	502	PR	21-JUN-2001; 2001EP-0000240.
DB	2604	GCAGCGTGTGAGTGTTCACCGTACTTTATTTGGATGGGATATTCAGCGGGTACAACT	2663	PR	21-JUN-2001; 2001EP-0000242.
QY	503	LysIleThrProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer	522	PR	21-JUN-2001; 2001EP-0000244.
DB	2664	CAGATTGGCCGGAATATAGCACCAACGAGGACATCCGAGGTACACACAGGATCGTTC	2723	PR	12-JUL-2001; 2001EP-0000280.
QY	523	ValProLeuProThrAlaThrSer	530	PR	30-JUL-2001; 2001EP-0000323.
DB	2724	GTGCGCACTGTCCAGCGGTGAGT	2747	PR	30-JUL-2001; 2001EP-0000327.
RESULT 4	ABZ78289			PR	02-AUG-2001; 2001EP-0000341.
ID	ABZ78289	standard; cDNA; 1872 BP.		PR	02-AUG-2001; 2001EP-0000342.
XX	XX			PR	02-AUG-2001; 2001EP-0000344.
XX	XX			PR	02-AUG-2001; 2001EP-0000344.
XX	XX			PR	09-AUG-2001; 2001EP-0000357.
DT	24-APR-2003	(first entry)		PR	16-AUG-2001; 2001EP-0000374.
DE	XX			PR	16-AUG-2001; 2001EP-0000377.
XX	XX			PR	20-SEP-2001; 2001EP-0000478.
XX	XX			PR	20-SEP-2001; 2001EP-0000552.
XX	XX			PR	22-OCT-2001; 2001EP-0000553.
XX	XX			PR	22-OCT-2001; 2001EP-0000554.
XX	XX			PR	22-OCT-2001; 2001EP-0000556.
XX	XX			PR	22-OCT-2001; 2001EP-0000557.
XX	XX			PR	15-NOV-2001; 2001EP-0000558.
XX	XX			PR	21-DEC-2001; 2001EP-0004464.
XX	XX			PR	21-DEC-2001; 2001EP-0005117.
OS	Aspergillus niger.			PS	(STAM) DSM NV.
XX	Key	Location/Qualifiers		XX	Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
XX	1..1872			XX	Klugbauer S, Wagner C, Fritz A, Von Gustadt W, Heinrich O;
XX	CDS	/*tag= a		XX	Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
XX	/*tag=			XX	Stiebler J, Albang R;
XX	/FC_number=	"3.4.16.6"		XX	WPI; 2002-723203/78.
XX				XX	P-PSDB; ABR38865.
XX				XX	Novel isolated protease polypeptide useful in laboratory, clinical,
XX				XX	pharmaceutical, chemical, diagnostic, personal care and industrial
XX				XX	applications -
XX				XX	Claim 1; Page 253-254; 394pp; English.
XX				XX	The invention relates to a novel isolated protease polypeptide. A
XX				XX	polypeptide or polynucleotide of the invention is useful for diagnosing a
XX				XX	fungal infection such as aspergillosis, or as a query sequence to perform
XX				XX	a search against public databases. A polypeptide of the invention is
XX				XX	useful in a selected number of industrial or pharmaceutical processes, in
XX				XX	laboratory or clinical processes, in food industry (baking, brewing,

Db 1629 AAGATCCATTGGATACGCGGCGATATCAATGGCTGTGTGGATTACTCGTCGAGC 1698
 QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyLeuLysAlaValAsnGluThr 279
 Db 1689 CCTTCGTCCTGAGCAGCGGTATAACAATACGTATGGATPCGAGGAAICAAATCGCACG 1748
 QY 280 ValTyrAsnTyrMetLysPheAlaAsnGluMetProAsnGlyCysGlnAspLeuIleSer 299
 Db 1749 CTCTACACCGCGGTATGGATAGTTGGAGCAAGCCITGGCGGTTCGAGGGATATGATCATC 1808
 QY 300 ThrCysLysGlnThr 1868
 Db 1809 GAGTGTGCGCATGCTGGCGAGCTCGGAGATCCCTCATGTATGCGACAATGAGACGGTA 1868
 QY 309 LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGly 328
 Db 1869 -----AATAGCACTCGGAGGCGCTCGGACACTCTTCGCGGGAGATCAAGAGC 1919
 QY 329 ProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro 348
 Db 1920 CTGTATACGAATACCTCGGCGGAGGAIACACGACATACGACATTCACGCCCGGATGCA 1979
 QY 349 ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGly 368
 Db 1980 GCTCTGTCGCTTACTTCGCGGCTTCTTGAATCGCCCATGGGTGCAAAAGGCACCTGGG 2039
 QY 369 ValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyrAlaPheGlnGlnThrGly 388
 Db 2040 GTCCCGGTGAACATATACCATGTCGTACAGGCGAGTGGGAACAGTTTCGCTCGAGCGGGC 2099
 QY 389 AspPheValTrpPro 403
 Db 2100 GATTAT-----CCGCGAAATGATCCCGGGAATGATCGGGATATTGGATACTTCCT 2153
 QY 404 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 423
 Db 2154 GACTCGGTGTCAAGTGCTATGGTATATGGGACCGGACTATGCTTGTCTGCGTGGCGC 2213
 QY 424 GlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 443
 Db 2214 GCGCGGGAAGATGTCAGCTGCTGTGGAGTACGAGGATCGGAGAGATTCGCTGCTGCT 2273
 QY 444 GlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGluTyr 461
 Db 2274 GGGTATGCGGAAGTCAGACAGCAAGTCA---TCCTACGTGGGGTCTAGTAGGCGAGTAT 2330
 QY 462 GlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnPro 481
 Db 2331 GGGAACTTCTGTTCCCGGTGCTTTTCAGGCGGCGCATGAGTGCCCATTTTATCAGCGC 2390
 QY 482 IleAlaSerLeuGlnPheAsnArgThrIlePheGlyTyrAspIleAlaGluGln 501
 Db 2391 GAAACGGGTATGAGATTTTAAATCGCGTCAGTTTAAATGGATATTCGCGGAGGCG 2450
 QY 502 LysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThr 519
 Db 2451 ATTTCTCTGGAGCAAGTACAGAGCTATGGAGCGGAGGACCGTCTCAACGTTGGATATC 2510
 QY 520 GlnSerSerValPro 524
 Db 2511 AAAAAAGAGTGGCG 2525

RESULT 6

ABZ78283

ID ABZ78283 standard; cDNA; 1665 BP.

XX AC ABZ78283;

XX DT 24-APR-2003 (first entry)

XX DE A. niger serine carboxypeptidase cDNA #2.

XX

Protease: fungal infection; aspergillois; food; tanning; detergent;
 protein solubility; viscosity; taste; texture; nutritional value;
 gene; ss.

Aspergillus niger.

Key Location/Qualifiers
 CDS 1..1665
 FT /*tag= a
 FT /EC_number= "3.4.16.6"

WO200268623-A2.

06-SEP-2002.

22-FEB-2002; 2002WO-EP01984.

23-FEB-2001; 2001EP-0200657.

23-FEB-2001; 2001EP-0200658.

23-FEB-2001; 2001EP-0200660.

26-FEB-2001; 2001EP-0200706.

26-FEB-2001; 2001EP-0200707.

26-FEB-2001; 2001EP-0200708.

26-FEB-2001; 2001EP-0200719.

28-MAR-2001; 2001EP-0000075.

28-MAR-2001; 2001EP-0000078.

28-MAR-2001; 2001EP-0000080.

28-MAR-2001; 2001EP-0000087.

28-MAR-2001; 2001EP-0000088.

21-MAY-2001; 2001EP-0000156.

21-MAY-2001; 2001EP-0000159.

21-MAY-2001; 2001EP-0000160.

21-MAY-2001; 2001EP-0000162.

21-MAY-2001; 2001EP-0000165.

21-MAY-2001; 2001EP-0000166.

21-MAY-2001; 2001EP-0000168.

21-JUN-2001; 2001EP-0000240.

21-JUN-2001; 2001EP-0000242.

21-JUN-2001; 2001EP-0000244.

21-JUN-2001; 2001EP-0000246.

12-JUL-2001; 2001EP-0000280.

12-JUL-2001; 2001EP-0000285.

30-JUL-2001; 2001EP-0000323.

02-AUG-2001; 2001EP-0000327.

02-AUG-2001; 2001EP-0000341.

02-AUG-2001; 2001EP-0000342.

02-AUG-2001; 2001EP-0000343.

09-AUG-2001; 2001EP-0000344.

16-AUG-2001; 2001EP-0000357.

16-AUG-2001; 2001EP-0000374.

16-AUG-2001; 2001EP-0000377.

20-SEP-2001; 2001EP-0000478.

20-SEP-2001; 2001EP-0000483.

22-OCT-2001; 2001EP-0000552.

22-OCT-2001; 2001EP-0000553.

22-OCT-2001; 2001EP-0000554.

22-OCT-2001; 2001EP-0000556.

22-OCT-2001; 2001EP-0000557.

15-NOV-2001; 2001EP-0000558.

21-DEC-2001; 2001EP-0000464.

21-DEC-2001; 2001EP-00005117.

(STAM) DSM NV.

PA

XX

PI

PI

PI

PI

XX

DR

XX

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 Stiebler J, Albarg R;

WPI; 2002-723203/78.

P-PSDB; ABR38859.

Novel isolated protease polypeptide useful in laboratory, clinical,

PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications -

claim 1: page 246-247; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform a search against public databases. A polypeptide of the invention is useful in a selected number of industrial or pharmaceutical processes, in laboratory or clinical processes, in food industry (baking, brewing, cheese manufacturing, meat tenderizing), in tanning industry and in the manufacture of biological detergents. A polypeptide may also be useful for improving protein solubility, extraction yields, viscosity or taste, texture, nutritional value, minimising of antigenicity or anti-nutritional factors, colour or functionality as well as processing aspects like filterability of the proteinaceous raw material. The sequences shown in AB278237-AB278293 represent cDNA encoding the A. niger proteases of the invention.

Sequence 1665 BP; 374 A; 455 C; 424 G; 412 T; 0 other; XX SO

Alignment Scores:

Alignment Scores:	
Pred. No.:	2.28e-83
Score:	941.00
Length:	1665
Matches:	209
Conservative:	72
Mismatches:	172
Indels:	90
Gaps:	13
DB:	24
Query Match:	32.59%
Best Local Similarity:	38.49%
Percent Similarity:	51.75%

US-09-712-338-2 COPY 19 555 (1-537) x ABZ78283 (1-1665)

Qy	13	GlnLeuProIysAsnProThrGlyValIysThrLeuThrThrAla-----AsnAsnVal	30
Db	49	CAATTTCCTCCGAGCGAAGGATCACTGCTCAAGTCCAGTTGCATGAGAATGTG	108
Qy	31	ThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyValIys	50
Db	109	ACTATTCTTTCAAGAGCCT-----GGAATTGCGAAACTACGCCGGGTCCGA	159
Qy	51	SerTyrSerGlyTyrValAspThrSerProGluSerHisThrPheThrPhePheGlu	70
Db	160	TCATTATCCGGGTATGTACACTTCCCCCGCCTCAACACAGCTCTTTTGGTTTTCGAA	211
Qy	71	AlaArgHisAsnProGluThrAlaProIleThrLeuThrLeuAsnGlyGlyProGlySer	90
Db	220	GCCCGCAAGATCCAGCAATCGCGCTCTGGCCATCTGCCTCAATGGCGGTCCGGTGC	279
Qy	91	AspSerLeuIleGlyLeuPheGluLeuGlyProCysHisValAsnSer---ThrPhe	109
Db	280	TGTCGCTCATGGGCTCCTTCAAGAATTAGTCTCTTTCATTGCATCAGACTCCAA	339
Qy	110	AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln	129
Db	340	ACCACAGTCTCAATCCCTGGAGTGTGAACATGAAGTCAATCTCTTATCTCTTCAC	399
Qy	130	ProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThr	149
Db	400	CCAATCAAGTCGGCTTCTCATACCATGTCCCAACAATGGCACTTGACA-----	450
Qy	150	GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla	169
Db	450	-----	450
Qy	170	ThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe	189
Db	451	-----GCTAAATGGGACTGCATTCGGGCTACGCTCTATGTCATTTCGCGCAACCTGG	504
Qy	190	LeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerIysAspPheSer	206
Db	505	TTTTTCGAGTCTCCACACTACAGCCCAACGATGATCGTCTC-----ACT	549
Qy	207	LeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlu	226

ID ABZ78241 standard; cDNA; 1581 BP.
 AC ABZ78241;
 XX
 XX
 XX 24-APR-2003 (first entry)
 XX
 DE A. niger serine carboxypeptidase cDNA #1.
 XX
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ss.
 XX
 OS Aspergillus niger.
 XX
 XX Key Location/Qualifiers
 FT 1..1581
 FT /*tag= a
 FT /EC_number= *3.4.16.6"
 XX
 PN W0200268623-A2.
 XX
 XX 06-SEP-2002.
 XX
 XX 22-FEB-2002; 2002W0-EP01984.
 XX
 PR 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 23-FEB-2001; 2001EP-0200659.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000162.
 PR 21-MAY-2001; 2001EP-0000165.
 PR 21-MAY-2001; 2001EP-0000166.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 15-NOV-2001; 2001EP-0000458.
 PR 21-DEC-2001; 2001EP-0000517.
 XX
 XX (STAM) DSM NV.
 XX
 XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;

PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albang R;
 XX
 DR WPI; 2002-723203/78.
 DR P-PSDB; AB338817.
 XX
 PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications
 XX
 PS Claim 1; Page 204-205; 394pp; English.
 XX
 CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
 CC A. niger proteases of the invention.
 XX
 SQ Sequence 1581 BP; 343 A; 473 C; 423 G; 342 T; 0 other;
 Alignment Scores:
 Pred. No.: 5.75e-81 Length: 1581
 Score: 916.50 Matches: 212
 Percent Similarity: 51.97% Conservative: 78
 Best Local Similarity: 37.99% Mismatches: 187
 Query Match: 31.75% Indels: 81
 DB: 24 Gaps: 13
 US-09-712-338-2_COPY_19_555 (1-537) x ABZ78241 (1-1581)
 QY 9 ValGlyArgArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrAlaAsn 28
 DB 46 GTCTCCGGGCCAGTTTGTGGCTCCGCCACCGATCATTCGCCACCAAGGATATCTC 105
 QY 29 AsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGly 48
 DB 106 GACATCCCGCTCCGCTACAAACAGTTC---CCACCGGCATTTGTGAGACTGATCCAGT 162
 QY 49 ValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPheThrPhe 68
 DB 163 GTCAAGAGCTTCCTCCGGTTACGTGATGTCGTCGAGCAGCAGCATCTCTTCIGTTC 222
 QY 69 PheGluAlaArgHis---AsnProGluThrAlaProIleThrLeuThrPheLeuAsnGly--- 86
 DB 223 TTCAGCGCGGCAACCAAGATCCACCGAGCTCCCTTGACCTGTGATCAATGAGGC 282
 QY 87 -----GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
 DB 283 ATGTCTGACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 342
 QY 103 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTyrAsnGluValSer 122
 DB 343 TCGGCATTTGAGCGCAATGGCTCCGCTACACACCCCTACTCTCTGGAACACGCCAGC 402
 QY 123 AsnLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
 DB 403 AACATGCTCTACATCAGCAGCCGCGGAGACCGGCTCTCTCTCTCTCTCTCTCTCT 450
 QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
 DB 451 -----ATTCCGGTTCCTCCGCTATGTTGATCTCTTCCACAGCAATGTTTATGGGC 501
 QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaAla 182

QY 469 lYrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAs 489
 Db 1982 GTAAGAGTGGACATGAGTTCCTCCCTTCATCAACCTTGCTGCGCTGGAGATTTGA 2041
 QY 489 nArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysLysIleTrpProSerTyrLy 509
 Db 2042 GCGCGTCATYGGCGGCAAGATGTGGCGACGGGAAAGATTCCTCGTCGAGTTTACA 2101
 QY 509 sThrAsnGlyThr----- 513
 Db 2102 GACGCTGGGCACGCCCAAGAGTTACTACCGGGAGGGCAACAGCAGCATTCAGTGGAGGT 2161
 QY 514 -----AlaThrAlaThrHis---ThrGlnSerSerValProLeuProThrAlaThrSe 530
 Db 2162 GTTGGATTCTCTGGCGACGTACACACACACACACACCATGCTCCGAACCCGGTGACCGGAG 2221
 QY 530 rMetSerSerValGlyMetAla 537
 Db 2222 GCTGAAGCGGATGGGACCGCT 2243

RESULT 9
 ABZ78226
 ID ABZ78226 standard; DNA; 3080 BP.
 XX AC ABZ78226;
 XX DT 24-APR-2003 (first entry)
 XX DE A. niger serine carboxypeptidase gene #2.
 XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW gene; ds.
 XX OS Aspergillus niger.
 XX PN WO200268623-A2.
 XX PD 06-SEP-2002.
 XX PF 22-FEB-2002; 2002WO-EP01984.
 XX PR 23-FEB-2001; 2001EP-0200657.
 XX PR 23-FEB-2001; 2001EP-0200658.
 XX PR 23-FEB-2001; 2001EP-0200660.
 XX PR 26-FEB-2001; 2001EP-0200706.
 XX PR 26-FEB-2001; 2001EP-0200707.
 XX PR 26-FEB-2001; 2001EP-0200708.
 XX PR 26-FEB-2001; 2001EP-0200719.
 XX PR 28-MAR-2001; 2001EP-0000075.
 XX PR 28-MAR-2001; 2001EP-0000078.
 XX PR 28-MAR-2001; 2001EP-0000080.
 XX PR 28-MAR-2001; 2001EP-0000087.
 XX PR 28-MAR-2001; 2001EP-0000088.
 XX PR 21-MAY-2001; 2001EP-0000156.
 XX PR 21-MAY-2001; 2001EP-0000159.
 XX PR 21-MAY-2001; 2001EP-0000160.
 XX PR 21-MAY-2001; 2001EP-0000162.
 XX PR 21-MAY-2001; 2001EP-0000165.
 XX PR 21-MAY-2001; 2001EP-0000166.
 XX PR 21-MAY-2001; 2001EP-0000168.
 XX PR 21-JUN-2001; 2001EP-0000240.
 XX PR 21-JUN-2001; 2001EP-0000242.
 XX PR 21-JUN-2001; 2001EP-0000244.
 XX PR 21-JUN-2001; 2001EP-0000246.
 XX PR 12-JUL-2001; 2001EP-0000280.
 XX PR 12-JUL-2001; 2001EP-0000285.
 XX PR 30-JUL-2001; 2001EP-0000323.
 XX PR 30-JUL-2001; 2001EP-0000327.
 XX PR 02-AUG-2001; 2001EP-0000341.
 XX PR 02-AUG-2001; 2001EP-0000342.
 XX PR 02-AUG-2001; 2001EP-0000343.

PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.
 PR 20-SEP-2001; 2001EP-0000478.
 PR 20-SEP-2001; 2001EP-0000483.
 PR 22-OCT-2001; 2001EP-0000552.
 PR 22-OCT-2001; 2001EP-0000553.
 PR 22-OCT-2001; 2001EP-0000554.
 PR 22-OCT-2001; 2001EP-0000556.
 PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 13-NOV-2001; 2001EP-0000464.
 PR 21-DEC-2001; 2001EP-00005117.
 XX XX (STAM) DSM NV.
 XX PA Edens L, Van Dijk AA, Krubasik P, Albertmann K, Stock A, Kimpel E;
 XX PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 XX PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 XX PI Stiebler J, Albang R;
 XX PS WPI: 2002-723203/78.
 XX DR P-PSDB: ABR38859.
 XX DR Novel isolated protease polypeptide useful in laboratory, clinical,
 XX PT pharmaceutical, chemical, diagnostic, personal care and industrial
 XX PT applications -
 XX PS Claim 1; Page 178-180; 394pp; English.
 XX CC The invention relates to a novel isolated protease polypeptide. A
 XX CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 XX CC fungal infection such as aspergillosis, or as a query sequence to perform
 XX CC a search against public databases. A polypeptide of the invention is
 XX CC useful in a selected number of industrial or pharmaceutical processes, in
 XX CC laboratory or clinical processes in food industry (baking, brewing,
 XX CC cheese manufacturing, meat tenderising), in tanning industry and in the
 XX CC manufacture of biological detergents. A polypeptide may also be useful
 XX CC for improving protein solubility, extraction yields, viscosity or taste,
 XX CC texture, nutritional value, minimising of antigenicity or
 XX CC anti-nutritional factors, colour or functionality as well as processing
 XX CC aspects like filterability of the proteinaceous raw material. The
 XX CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
 XX CC A. niger proteases of the invention.
 XX SQ Sequence 3080 BP; 755 A; 803 C; 735 G; 787 T; 0 other;

Alignment Scores:
 Pred. No.: 1,11e-77 Length: 3080
 Score: 887.50 Matches: 223
 Percent Similarity: 48.15% Conservative: 77
 Best Local Similarity: 35.79% Mismatches: 191
 Query Match: 30.74% Indels: 133
 DB: 24 Gaps: 17

US-09-712-338-2_COPY_19_555 (1-537) x ABZ78226 (1-3080)

QY 13 GlnLeuProLysAsnProThrGlyValLysThrLeuThrAla-----AsnAsnVal 30
 Db 563 CAATTTCTCCCGAGCGGAGGCATCACTGTGCTCAAGTCCAGTTGCATGAGAATGTG 622
 QY 31 ThrIleArgTyrLysGluProGlyAlaGlu----- 40
 Db 623 ACTATTCTTTCAAGAGGCTGTGTGAGA-GTATCTAGAAATAGCTTTTATCTCGATGC 681
 QY 41 -----GlyValCysGluThrThrProGlyValLysSerTyrSergly 54
 Db 682 CGTGTGATTTGTCAGCTCGAATTTGCGAACTACGCCGGGTGTCGATCTTATTCGGGC 741
 QY 55 TyrValAspThrSerPro----- 60
 Db 742 TATGTACACCTTCCCCCGGTTTCCTTCCGAGGACAGAGAGAGGAGGATTATCTT 801

QY 61 -----Glu-SerHisTh 64
 Db 802 ATCAACAGTAAGCAATCTCGAACAATTCGAGCATGAGCAATTCACGAGCTCAACAG 861
 QY 64 rPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLe 84
 Db 862 CTTCTTTGGTTTTCGAAGCCGCAAGATCCAGCAATCGCGCTCTGGCCACTGGCT 921
 QY 84 uAsnGlyClyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHi 104
 Db 922 CAATGGCGGTCGGGTGGCTCGCTCGCTCATGGGCTCTTGAAGAATAGTCTGTGTC 981
 QY 104 sValAsnSer----rPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAs 123
 Db 982 CATTCGATCAGACTCCAGACCAACAGCTCTCAATCTTGGAGTGGACAAATGAGTCAA 1041
 QY 123 nLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSer 138
 Db 1042 TCTTCTATTCTTTCACAGCCCACTCAAGTCGGCTTCTCATACGATGTCCCAACAATGG 1101
 QY 139 -----AspThrValAspGlySerIleAsnProValThrGly-----Va 151
 Db 1102 CACTTTGGTTCGGACTCGGACGCGAAGAGATAGTTTCGGGTGATTTCTCCATGA 1161
 QY 151 lValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLe 171
 Db 1162 TGTTCGCCAGTCCAACTTCACCCATCATGTGTGTACCTTTGCAAGC-----CAGAAGCT 1215
 QY 171 uIleAspThrAsn-----LeuAlaGluAlaAlaTrpGluIleLeuGlnG 188
 Db 1216 TGCACAGACAGCTAAATGGGACTCGAATCGCGGCTCAGGCTCTATGGCATTTTCGGCAAC 1275
 QY 188 yPheLeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerLysAspPh 205
 Db 1276 CTGGTTCGAGTTCACACATACACAGCAAGCCAGATGATCGTGC----- 1321
 QY 205 eSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTy 225
 Db 1322 -AGTCTCTGGCTGAAATTCAGGAGCCATATGTCAGGCACTTTTCGGTTCCTCCA 1380
 QY 225 rGluGluAsnGluArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPh 243
 Db 1381 ACAGCAGAAATGACAAAATCGCAGAGGGAGCTGCAGAAAGAGCGTGACAGTATTTGCATCT 1440
 QY 243 eAsnSerLeuGlyIleIleAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyrPr 263
 Db 1441 CGACAGCTTGGCATTTGTAACGGCTTGATGGATATGGTATCCAAAGAGAGGCTTACAT 1500
 QY 263 oGluPheAlaValAsnAsn----- 269
 Db 1501 TACTTGGCCATACATAACGTAAAGCTCGCCCTTCCTTCATTCACCTCGGTAAATGCCVA 1560
 QY 270 -----ThrTyrGlyIleLeuAlaValAsnGluThrValTyrAsnTyrMetLysP 286
 Db 1561 ATTCAAGTTCAGACCTACGCGCTCGAAATCTTCGATAAACCCCTCTACGAAGAAGTGTATGT 1620
 QY 286 heAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnA 306
 Db 1621 ATAACCTGGACCATCCAGGAGC----- 1643
 QY 306 rGThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp--- 324
 Db 1644 -----TTTCGATCAGGCGCTCGCTCGCAAGCGGCTTGAAGAAGACGGATTCOG 1695
 QY 325 -----AsnVal----- 326
 Db 1696 GCTTGCTCCTACTCAGGGAAGATATCTGAAATTTGCGAGGCGCTTGCACATGAGATGG 1755
 QY 327 --GluGlyPro-----TyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisP 344
 Db 1756 GAGATGGCCCATACCTACTACACACCTTCAATCGCGGGTGTACGACATCGGCCCATC 1815

QY 344 rOIYrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValM 364
 Db 1816 CTAGAAGACGCCATTCCTCTCCCAAGCATCTCGGATATTTGACGAGGAGTCCGTC 1875
 QY 364 eAspAlaIleGlyValAsnIleAsnTyrThrGluSerAsnAsnAspValTyrIYrAlaP 384
 Db 1876 TTGCGGCTCTGGGTACAGTCAATTCACAICTCTTCGATCGCGTGGTGGTCAACAGT 1935
 QY 384 heGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIleLeuA 404
 Db 1936 TCATAAAAACCTTTGATATCGTCCACGCGGCTTCCTGTGATGCAATTGCTACCTCC 1995
 QY 404 laLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheG 424
 Db 1996 ACATGGTGTATAAAGTACACATGATGACGAGATCGTGATTACGCTGCAATGGTGG 2055
 QY 424 lYGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaG 444
 Db 2056 GGGGGAANAACCCAGCTTCGAGTTCGATATTCGCTATTCACGGAATTCGCCGACAGG 2115
 QY 444 lyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnP 464
 Db 2116 GATATCTCCCATCTCTACGCGGATCAGCGGATCAGCGGATCAGCGGATTCGGCAACT 2175
 QY 464 heSerPheThrArgValIYrGluAlaGlyHisGluValProTyrTyrGlnProIleAla 484
 Db 2176 ACAGCTTCACTCGCGCTCTTCCAAAGCGGCAAGGTCCTCTTACCAGCTGTGGGG 2235
 QY 484 erLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGlyGlnLysLysI 504
 Db 2236 CGIATGAGATCTTCATCGGCGGACATTCACAAGAATATCCCTACTGGGCTCTTGGGTG 2295
 QY 504 leTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValP 524
 Db 2296 TTGATGACGAATTCAGTCGGTGGACCTAAGGATACGTGSCATATCAAGAATATCCCTC 2355
 QY 524 rOleu 525
 Db 2356 CTATT 2360
 RESULT 10
 ABZ51997
 ID ABZ51997 standard; cDNA; 1007 BP.
 XX
 AC ABZ51997;
 XX
 DI 28-MAR-2003 (first entry)
 XX
 DE Aspergillus oryzae polynucleotide SEQ ID NO 1110.
 XX
 KW Aspergillus oryzae; fermentation; fungus; industrial; EST;
 KW -expressed sequence tag; gene; ss.
 OS
 XX Aspergillus oryzae.
 PN W0200279476-A1.
 XX
 PD 10-OCT-2002.
 XX
 PF 22-MAR-2002; 2002WO-1500890.
 XX
 PR 30-MAR-2001; 2001JP-0098371.
 XX
 PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (NARE-) NAT RES INST BREWING.
 PA (NORQ) NAT FOOD RES INST MIN AGRIC.
 XX
 PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
 PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
 DR WPI; 2003-046817/04.
 XX
 PT Detection of expression of specific Aspergillus genes for monitoring

the fermentation and growth conditions of the fungus, using DNA probes

Claim 1: SEO ID NO 1110: 48pp + Sequence Listing; Japanese.

The invention relates to a polynucleotide having any of 6006 specific sequences (ARZ50888-ARZ56893), which are expressed by a fungus under specific culture conditions including one or more of eutrophic, oligotrophic, solid, early germination, alkaline, high temperature, low temperature or maltose culture or polynucleotides stringently hybridising to these sequences. The polynucleotides are useful for monitoring the progress of fermentation and the growth conditions of a fungus, especially of *Aspergillus oryzae* which is widely used in industrial fermentation. Also monitoring for fungal contamination.

CC Recommendation: Also monitoring for further developments.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](ftp:wipo.int/pub/published_pct_sequences).

Sequence 1007 BP; 253 A; 270 C; 257 G; 224 T; 3 other;

Alignment Scores:	3.42e-60	Length:	1007
Pred. No.:	704.50	Matches:	156
Score:	61.36%	Conservative:	52
Percent Similarity:	46.02%	Mismatches:	97
Best Local Similarity:	24.40%	Indels:	36
Query Match:	25	Gaps:	9
DB:			

US-09-712-338-2 COPY 19 555 (1-537) X ABZ51997 (1-1007)

197	Qy	AspSerArgValGlnSerLysAspPheSerLeuThrGluSerTyrGlyGlyHisIstyr	211
18	Db	GATCACCAGGATP-----AGCATCTGGAGGAGC--TACGGAGGACGGTAC	60
217	Qy	GlyProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerVal	236
61	Db	GGGCCAICTTTCACTGCTTTCTTCAGGAGCAGAAIAGAGAAGATTGCCAACSS-TCCATT	119
237	Qy	-----AsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIle	253
120	Db	GACATPCGACGATGCTCACTATTATCACTCGGACACCCCTCGGAATTATCAATGGCTGGCTC	179
254	Qy	AspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIle	273
180	Db	GATCTCCTGGTGCAGTGGCCCATCATUCCCGCAGATAGCATACACAATACCTATGCCATT	239
274	Qy	LysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGly	293
240	Db	GAGGCTATCAACAAACCTGTATATGATATGGCGATGGAGGCTTGGAGCAAGCCCGGAGCC	299
294	Qy	CysGlnAspLeuIleSerThrCysLysGln-----AsnPheIleGlu	303
300	Db	TGCAAGGATGACATCATCGAATGTGCTCGGTGGCTGCAGAGGGAGACCCCGAAATGTAC	359
304	Qy	--ThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCys	322
360	Db	GGCAACAATGACACCGTC-----AACAGGCTGTGCCGAAGCAACAACACTACTGC	410
323	Qy	ArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg	342
411	Db	AGCAACCAAGTAGAAGGCCCATATTCACTGTACTCTGGCAGGAGCTATTACGACACTCG	470
343	Qy	HisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSer	362
471	Db	CAMTTTCGACCCCGATCCCTTCCCCCTCCGCTACTTTGGTTTCTTAACCAAGCACTGG	530
363	Qy	ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspValTyrTyr	382
531	Db	GTCCAAAGAGCCCTCGGGGTTCGGTGTATTTCACCGGATCTGTGTATAGCTCTACAT	590
383	Qy	AlaPheGlnGlnThrGlyAspPheValTyrPro-----AsnPheIleGlu	397
591	Db	GGCTTTTCCGCAACAGGTGACTAC-----CCAGCTCCGATGTACCGGGTACCTGGAG	644

CC vasotropic activity and can be used in vaccines or for gene therapy. The
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
 CC antisense molecules and antibodies are useful as medicaments or in
 CC preparing a medicament for treating, preventing and/or alleviating
 CC diseases associated with yeast or fungi or proliferative disorders, such
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds
 CC or polypeptides, or the genetically modified organism are useful for
 CC preparing a medicament for modifying the endogenous flora of humans and
 CC other mammals. The vaccine is useful for immunising against yeast or
 CC fungal infections. Apoptosis-related diseases include autoimmune disease,
 CC ischaemia, diseases related with viral infections or neurodegenerations.
 CC This sequence represents a polynucleotide associated with the Bax gene
 CC described in the disclosure of the invention.

XX
 SQ sequence 2027 BP; 583 A; 380 C; 427 G; 637 T; 0 other;

Alignment Scores:

Pred. No.: 5.76e-53 Length: 2027
 Score: 636.00 Matches: 167
 Percent Similarity: 46.29% Conservative: 70
 Best Local Similarity: 32.62% Mismatches: 197
 Query Match: 22.03% Indels: 78
 DB: 24 Gaps: 14

US-09-712-338-2_COPY_19_555 (1-537) x ABQ76315 (1-2027)

QY 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeu-----ProLys 16
 DB 627 CTTCCAGAAATACACGAAACCTTAAATGGACCGTTTGAATCAGCATGATCCCGCTG 686
 QY 17 AsnProThrGlyValLysThrLeuThrThraAlaAsnAsnValThrIleArgTyrLysGlu 36
 DB 687 TTTTACAACCTTTATTTCTCTGTGGACACAGATTACAGTTTGACATTAGACAGTAGAT 746
 QY 37 ProGlyAlaGluGlyValCysGluThrProGlyValLysSerTyrSerGlyTyrVal 56
 DB 747 CTTCTTAAACTAGGAATT-----GACACCCGTAATAAATGGICGGGTATACATG 794
 QY 57 AspThrSerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGlu 76
 DB 795 GACTATAAGGATTCACAAACACTTTTITACTGGTTTGAAGTAGGACGATCCGCT 854
 QY 77 ThrAlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 96
 DB 855 AAGCAGCCAAATATCTTGTAAATGGTGACCTGGTTGTTCCICGTTTACTGGGTG 914
 QY 97 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspSerTyrIleAsnProHis 116
 DB 915 CTATTTGAACCTAGGCCCTCATCAATGGCGCGGATATGAACCAATCCCAATCCCTAT 974
 QY 117 SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 136
 DB 975 TCTTGGAAATAATACCGTCAATGATCTCTTAGACAGCCACTCGGAGTGGCTTTTC 1034
 QY 137 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSer 156
 DB 1035 TATGGTGAT----- 1043
 QY 157 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsn 176
 DB 1044 -----GAAAAAGTCTCTCTACAAAA 1064
 QY 177 LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 196
 DB 1065 TTACAGGCAAGATCGGTACATTTTCCGGAAATGTTTTCGAAGCTTTTCTCATTTA 1124
 QY 197 AspSerArgValGlnSerLysAspPheSerLeuThrThrGluSerTyrGlyGlyHisTyr 216
 DB 1125 -----CGCTCCCAAGATTTCCCAATTCAGGCGCAATCCCTATCCAGACATAT 1172
 QY 217 GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer 235
 DB 1173 ATCCCTCAAATTGCATGATCGTTGTCAAGAACCCCTGAAGA----- 1217

QY 236 ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGlu 255
 DB 1218 -----ACGTTCAATTTAAGTTTCAATGATGTTAATGGTATGATACACACCT 1265
 QY 256 AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 274
 DB 1266 TTGATTCAGCAGATATTTATCAACCAATGGCATCGGGAAAGGGGGGTATACACCTGTT 1325
 QY 275 AlaValAsnGluIleValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 294
 DB 1326 CTCCTCATCAGAAGAATGTGAGAAAATGATTAAGCTGCAGCTCGI----- 1370
 QY 295 GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu 314
 DB 1371 -----TGCTAGGTGTAACAAAGTTATGTTATGCTTCTAAATCAAGT 1412
 QY 315 -----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla 332
 DB 1413 TTACATGCAATAGTCGCCACCTGCTTACTGTGACTCTGACCTTTTGGAAACCGTACATTAAC 1472
 QY 333 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 348
 DB 1473 ---ACAGGACTCAACGCTCTATGACATAGAGGCCCTCTGAGATAATAGTACTAGTATGTT 1529
 QY 349 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 365
 DB 1530 ATGTGTTATACAGGTCTCCGCTATGTCAGCAGTATGAAATTTTCTGAAAGTTCAAGAA 1589
 QY 366 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPhe 384
 DB 1590 ACGTAGGTCGCGAGTCATATATTTCTGGCTGTGATAATGACGTTTCCACCGGATTT 1649
 QY 385 GlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAla 404
 DB 1650 ITGTTTACGGCGCATGGAAGTAAACCA--TTTCAACAATATATGCTGAATTAATAAT 1706
 QY 405 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGly 424
 DB 1707 CACAACATTCGGTATTATTAATATATGCGGTGATAAGGATATATTTCTAATTTGGTGG 1766
 QY 425 GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly 444
 DB 1767 AACCATGCTTGGTCCATGAGTTGGAATGGATCAATAAACAGTATGATCAGAGAAGGATG 1826
 QY 445 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462
 DB 1827 TTAAGACCATGGGTTCAGTAAAGAAACAGSTCAAGAGTTGGACAAAGTCAAGAACTATGCC 1886
 QY 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482
 DB 1887 CCTTTCACCTTTTGTAGAATAACGATGCGGTGATATGTTGTCCTATGATCAACCGGAG 1946
 QY 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 494
 DB 1947 GCAAGTTTGAATGGTCAACAGTTGGATTTCCGGT 1982
 RESULT 12
 ABZ78243
 ID ABZ78243 standard; cDNA; 1611 BP.
 XX
 AC ABZ78243;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE A. niger carboxypeptidase Y cDNA #1.
 XX
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 XX gene; ss.
 OS Aspergillus niger.
 XX

Db	872	-----GTACAGCACACCGTTGCTGGCAAGACCTCTATGCTTG 913	XX	AAT28283;
QY	186	LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205	AC	14-AUG-1996 (first entry)
Db	914	CTTACCTCTTCTTCAAAATTCGCC-----GAGTATGCCAAGCAGGACTTC 961	DT	A. niger Bo-1 genomic carboxypeptidase Y clone.
QY	206	SerLeuThrThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyr 225	DE	Carboxypeptidase Y; CPY; ascomycete; deuteromycete; host cell;
Db	962	CACATTCGGGTGAACCTATGCTGTCTACTATATCCCGCTCTTCTCTCGGAGATTGG 1021	KW	protease deficiency; ds.
QY	226	GluGlnAsnGluAlaGlnLeuAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245	XX	Aspergillus niger strain Bo-1.
Db	1022	TCTCACAGAGAGCG-----AACATCAACCTGCAGCTCC 1054	XX	Key
QY	246	LeuGlyIleLeuAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyrProGluPhe 265	XX	Location/Qualifiers
Db	1055	GTTCTTATTTGGCAACGGTCTACCGGCTCTCACTGAGTACGAGTACTACCGTCCGATG 1114	FT	1..571
QY	266	AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 285	FT	/*tag= a
Db	1115	GCCTGTGGTACGGT---GGTACCAGCTGCTTGGACGAG---GGCTCCTGCCAGGCC 1168	FT	/codon_start= 64
QY	286	PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 305	FT	572..633
Db	1169	ATGGACACAGCCCTTCTCTCGC---TGCCAGTCTATGATTGAGTGTGTGTAAGTCCGAG 1225	FT	/*tag= b
QY	306	ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 325	FT	634..2068
Db	1226	AGCGCT-----TGGGTTTGTCGCCGCTCCATCTACTGTACACAGCC 1270	FT	/*tag= c
QY	326	ValGluGlyProTyrTyrAlaPheAlaGlyValTyrAspIleArgHisProTyr 345	XX	WO9609397-A1.
Db	1271	CTCCTGCCCTTACACGCG---ACGGACAGACGCTACGATGTTCTGGTGAAGTGC 1327	XX	28-MAR-1996.
QY	346	AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360	XX	19-SEP-1995; 95WO-US11945.
Db	1328	GAGGATAGCTCCAACTCTCTACTCGGCCATGGCTACGTACGCGACTACCTGAACAAG 1387	XX	20-SEP-1994; 94US-0309341.
QY	361	AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAspAsp 379	XX	(NOVO) NOVO NORDISK BIOTECH INC.
Db	1388	ACCGAGTCAITGAGGCTGTGGCGTGAAGTCAACGGCTACGCTGCACCTTGAC 1447	XX	Thompson SA, Yaver DS;
QY	380	ValTyrTyrAlaPheGlnGlnThrGlyAspPheValIrrPro-----AsnPheIleGlu 397	XX	WPI: 1996-188458/19.
Db	1448	ATCAACGGCAACTCTCTTCCAGGGTACTGGATGAAGCCCTACCCACCGTCTCGTTCCG 1507	XX	P-PSDB; AAR96737.
QY	398	AspLeuGluGluLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416	XX	Nucleic acid construct encoding a filamentous ascomycete or
Db	1508	GGACTCTCTGGAG-----CAGATCCCTGTC-----CTGATCTACGCTGTGTGCGCC 1552	XX	deuteromycete carboxypeptidase Y - useful to produce host cells
QY	417	AspTyrIleCysAsnTyrPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGln 436	XX	modified to produce reduced amounts of carboxypeptidase
Db	1553	GATTCATCAGCACTGGCTGGGCAACAGCGCTGAGCTGAGCCCTTGACTGCCCGGA 1612	XX	Claim 4; Fig 1A-E; 46pp; English.
QY	437	AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451	XX	The CPY gene (T28283) of Aspergillus niger strain Bo-1 codes for
Db	1613	CAGGCTCAATATGCTCCGCTAAGCTGAGGACCTGGCTGCTGCTGCTGCTGCTGCTGCT 1672	CC	carboxypeptidase Y pre-propeptide (R96737). It was obt'd. by PCR
QY	452	GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471	CC	amplification of A. niger genomic DNA using degenerate primers (see
Db	1673	GGCAAGAGATCGGCCAGGTCAGTCCATCGGCAACTTCACCTTCATGCGCTCTATGGC 1732	CC	also T28285-92) based on known carboxypeptidase sequences, and use
QY	472	AlaGlyHisGluValProTyrTyrGlnPheAlaSerLeuGlnLeuPheAsnArgThr 491	CC	of a PCR product to screen a genomic DNA bank of A. niger in EMBL4.
Db	1733	GGTGGCCACATGTTCCCGATGGACACCCGAGTCGAGTCTTGATATCTTCAACCGCTGG 1792	CC	The gene is useful in the creation of CPY-deficient mutants of
QY	492	IlePheGly---Tyr 495	CC	Aspergillus. For example, a selectable marker may be cloned into the
Db	1793	TTGGGAGGTGAATGG 1807	CC	middle of the CPY gene. The resulting CPY-deficient mutant is useful
Db	1793	TTGGGAGGTGAATGG 1807	CC	as a host for prodn. of heterologous protein.
Db	1793	TTGGGAGGTGAATGG 1807	XX	Sequence 2068 BP; 436 A; 601 C; 522 G; 509 T; 0 other;
Db	1793	TTGGGAGGTGAATGG 1807	XX	Alignment Scores:
Db	1793	TTGGGAGGTGAATGG 1807	XX	Pred. NO.: 2.15e-47 Length: 2068
Db	1793	TTGGGAGGTGAATGG 1807	XX	Score: 580.00 Matches: 154
Db	1793	TTGGGAGGTGAATGG 1807	XX	Percent Similarity: 46.25% Conservative: 80
Db	1793	TTGGGAGGTGAATGG 1807	XX	Best Local Similarity: 30.43% Mismatches: 175
Db	1793	TTGGGAGGTGAATGG 1807	XX	Query Match: 20.09% Idels: 98
Db	1793	TTGGGAGGTGAATGG 1807	XX	DB: 17 Gaps: 18
Db	1793	TTGGGAGGTGAATGG 1807	XX	US-09-712-338-2_COPY_19_555 (1-537) x AAT28283 (1-2068)
Db	1793	TTGGGAGGTGAATGG 1807	XX	QY 27 AlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db	1793	TTGGGAGGTGAATGG 1807	XX	DB 457 GCCTATGATCTCAGGTCACAGAACCCATCTGCTCTCTTGGATC-----GAC 507
Db	1793	TTGGGAGGTGAATGG 1807	XX	QY 47 ProGlyValLysSerTyrSerGlyTyrValAsp----- 57
Db	1793	TTGGGAGGTGAATGG 1807	XX	DB 508 CCGCGGTGAAGCAGTACACCGGTATCTCGATGACACAGAGATGATAGCATTTGTTTC 567
Db	1793	TTGGGAGGTGAATGG 1807	XX	QY 58 -----ThrSerProGluSerHisThrPheTrp----- 67

```

Db 568 TACTCTAAGCACACACCTTGGTTCAAGATCACGCTTTTATAT-GCTCTGATATCTAACGC 626
QY 68 -----PheGluAlaArgHisAsnProGlnThrAlaProIleThrLeuTrpLeu 84
Db 627 AACTTAGGTTCTTCGAGTCTCGCAATGACCCGAGATGATCCCGTTGTTCTTGGGTCG 686
QY 85 AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuLeuGlyProCysHis 104
Db 687 AACGGTGGCCCTGGGTCTCTCCCTCACCGGTCTCTCATGAGCTTGGCCCTAGCAGC 746
QY 105 ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 124
Db 747 ATCAACAAGAAGATCCACCGGCTCTACAATGACTACGCTTGAACATCCCAACGCGTCG 806
QY 125 LeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer 144
Db 807 ATCTTCCTTGACCAACGCTGCAATGCTACTCTACAGTAACCTGCT----- 857
QY 145 IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr 164
Db 857 ----- 857
QY 165 ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGlu 184
Db 858 -----GTCAGCGACACGCTCGCTGCTGGCAAGGACGCTCTATGCC 896
QY 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 204
Db 897 TTGCTTACCCCTCTTCAACAATTCCTC-----GAGTATGCTAAGCAGGAC 944
QY 205 PheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPhe 224
Db 945 TTCACATATGCGGTGAATCTTATGCTGCTCACTATATCCCGCTTCGCTCGGAGATC 1004
QY 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244
Db 1005 CTGCTCACAGAAGACGC-----AACATCAACCTGCAG 1037
QY 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGlu 264
Db 1038 TCCGTTCTCATGGCACGGTCTACCGACGGATACACCCAGTACGAGTACTACCGTCCC 1097
QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
Db 1098 ATGCGCTGCGGTGACGCGGTATCCAGCTCTCTTGGACGAGAGCTCTGCCAGTCCAIG 1157
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db 1158 GAC-----AACGCTCTCTCTCGC---TGCCAGTCTATGATGAGTCTGTGTACAGTTC 1208
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 324
Db 1209 GAGAGCGCT-----TGCGTTGTCTCCGCGCTCCACTCTACTGTAAACAC 1253
QY 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344
Db 1254 GCCTCTCTTGCCTTACCAGCGC---ACTGGCAGAACGCTATGATGCTCCGGTGAAG 1310
QY 345 TyrAspAspPro-----ThrProProSerTyrTyrAsnLysPheLeuAla 359
Db 1311 TCGGAGGATAGCTCTAACCTTTGCTACTCGGCTACGCTACGCTACGCTACCTGAAC 1370
QY 360 LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn 378
Db 1371 AAGCCCAAGTCAATGAGGCTGTGGCGCTGAGGCTCAACGGCTACGACTCGTGCACACTT 1430
QY 379 AspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIle 396
Db 1431 GACATCAACGCAACTCTCTCTTCCAGGCTGATGATGAGGCTTACCAACCGCTCGGTI 1490
QY 397 GluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp 415

```

```

Db 1491 CCGGGACTCCTGGAG-----CAGATCCCGTGC-----TTGATCTATGCGCGTGAT 1535
QY 416 AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSer 435
Db 1536 GCTGATTTTCATTTGCAACTGGCTGGGCAACAGGCTGGACCTGGAGTGGAGTGGCC 1595
QY 436 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal----- 450
Db 1596 GGACAGGCTGAATATGCCCTCCGCTGAGCTGGAGGATCTGCTCATTTGCGACAATGAGC 1655
QY 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470
Db 1656 ACGGCAAGCAAGATGGCCAGGTAAAGTCCCATGGCAACTTCACTTCAATGCTCTCTAT 1715
QY 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490
Db 1716 GGTGTGGCCACATGCTCCGATGGACCAAGCCGAGTCGAGTCTGAGTTCCTCAACCG 1775
QY 491 ThrIlePheGly---Trp 495
Db 1776 TGGTTGGGAGGTGAATGG 1793

RESULT 15
ABZ78186
ID ABZ78186 standard; DNA; 2660 BP.
XX
AC ABZ78186;
XX
DT 24-APR-2003 (first entry)
XX
DE A. niger carboxypeptidase Y gene #1.
XX
KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW gene; ds.
XX
OS Aspergillus niger.
XX
PN WO200268623-A2.
XX
FD 06-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-EP01984.
XX
PR 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.

```



```
Qy 436 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal----- 450
Db |||:|:|:|:| ||| ||| ||| :|:|
2029 GGACAGGCTGAATATGCTCGCTGAGCTGGAGGATCTGGTCATTGTCGACAATGAGCAC 2088

Qy 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470
Db ||| :|:| :|:|:|:| :|:| :|:|:|:|:|:|:|:|:|:| |||:|:|:|:|
2089 ACGGCAAGAAGATTGCCAGGTTAGTCCCATGGCAACTTCACCTTCATCGCTCTCTAT 2148

Qy 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2149 GGTGGTGGCCACATGGTCCCGATGGACCGCCGAGTCGAGTCTCGASTTCTTCAACCGC 2208

Qy 491 ThrIlePheGly---trp 495
Db :|:| :|:| |||
2209 TGGTTGGGAGGTGAATGG 2226
```

Search completed: September 16, 2003, 23:02:52
Job time : 499 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:02:15 ; Search time 554 Seconds
(without alignments)
7835.192 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662

Perfect score: 1608

Sequence:

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL1/cgcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
1	1608	100.0	1662	19	AAV28620		A. oryzae ATCC2038

ALIGNMENTS

RESULT 1
AAV28620

ID	AAV28620	standard; DNA; 1562 BP.
XX		
XX	AAV28620;	
AC		
XX	27-AUG-1998	(first entry)
DT		
XX		
DE	A. oryzae ATCC20386	carboxypeptidase I DNA.
XX		
KW	Carboxypeptidase I;	flavour improving agent; hydrolysate; proteinaceous;
KW	food industry; ss.	
XX		
OS	Aspergillus oryzae.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1..1562
FT		*tag= a
FT		/product= carboxypeptidase I
FT		/note= "partial coding sequence"
XX		
PN	WC9814599-A1.	
XX		
PD	09-APR-1998.	
XX		
PF	03-OCT-1997;	97WO-US17977.
XX		
PR	27-NOV-1996;	96US-0757534.
PR	04-OCT-1996;	96US-0726880.
XX		
PA	(NOVO) NOVO NORDISK BIOTECH INC.	
PA	(NOVO) NOVO-NORDISK AS.	
XX		
PI	Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;	
PI	Klotz A, Mathisen TE, Rey M;	
XX		
DR	WPI; 1998-240098/21.	
DR	P-PSDB; AAW56099.	
XX		
PT	Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous	
PT	substrates, useful for improving flavour of foods	
XX		
PS	Claim 2; Fig 3; 82pp; English.	
XX		
CC	This DNA sequence encodes carboxypeptidase I from Aspergillus oryzae.	
CC	This polypeptide has an optimal activity in the range of pH 3.0-7.5 at	
CC	25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a	
CC	residual activity of at least 65.5% after 30 minutes at pH 4.0 and	
CC	60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where	
CC	N-CBZ is N-carboxy and X is any amino acid. The carboxypeptidases	
CC	can be used for obtaining hydrolysates (which can be enriched in free	
CC	glutamic acid or peptide bound glutamic acid residues) from proteinaceous	
CC	substrates. The carboxypeptidases can be used in flavour-improving	
CC	compositions in the food industry. The products can also be used for the	
CC	production of polypeptides free of carboxypeptidase activity.	
XX		
SO	Sequence 1562 BP; 396 A; 468 C; 398 G; 400 T; 0 other:	

[illegible]

Db 235 GAGTCCCATACCTCTCTCTGTTCTTGAAGCCAGACATACCCAGAACTGCACTATC 294
QY 241 ACATTGGTGGTTGAATGTGGCCCTGGAGGAGATCTTTGATCGGTCTCTTCGAGACGTTG 300
Db 295 ACATTGGTGGTTGAATGTGGCCCTGGAGGAGATCTTTGATCGGTCTCTTCGAGAGATTG 354
QY 301 GGCCCTTGGCATGCAATTCGACTTTTGGATGACATACATCAACCCCTCACTCGTGAAGAG 360
Db 355 GGCCCTTGGCATGCAATTCGACTTTTGGATGACATACATCAACCCCTCACTCGTGAAGAG 414
QY 361 GTCTCCAAATTACTATTCTTCGTCGCCAGCCATTGGAGTGGCTTTTTCATATGATGATACG 420
Db 415 GTCTCCAAATTACTATTCTTCGTCGCCAGCCATTGGAGTGGCTTTTTCATATGATGATACG 474
QY 421 GTTCATGGTCCCAATTACCCCTGTAACCTGGGTCTGCAAAATTCGAGCTTTGCGAGGTT 480
Db 475 GTTCATGGTCCCAATTACCCCTGTAACCTGGGTCTGCAAAATTCGAGCTTTGCGAGGTT 534
QY 481 CAGGCGCGGTACCCAAACCATTTGATGCCACCTGTGATCGATACCAATCTTGGCGCAGAG 540
Db 535 CAGGCGCGGTACCCAAACCATTTGATGCCACCTGTGATCGATACCAATCTTGGCGCAGAG 594
QY 541 GCCCTTTGGGAGATCCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGTG 600
Db 595 GCCCTTTGGGAGATCCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGTG 654
QY 601 CAGTCTAAGGACTTCAGTCTATGACGAGAGCTATGGAGGCACTATGGTCTCTGCAATTC 660
Db 655 CAGTCTAAGGACTTCAGTCTATGACGAGAGCTATGGAGGCACTATGGTCTCTGCAATTC 714
QY 661 TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTAAATGGTGTTCAG 720
Db 715 TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTAAATGGTGTTCAG 774
QY 721 CTTAATTTCACTCTCTGGGAATTTAAGCGGATCATCGAGGGGATCCAGGCCCTT 780
Db 775 CTTAATTTCACTCTCTGGGAATTTAAGCGGATCATCGAGGGGATCCAGGCCCTT 834
QY 781 TACTACCTGAATTCGCTGTGCAACAATACCTACGGTATCAAGGCTGTCAACGAGACGGTC 840
Db 835 TACTACCTGAATTCGCTGTGCAACAATACCTACGGTATCAAGGCTGTCAACGAGACGGTC 894
QY 841 TACAACCTACATGAAGTTTGCCAAACCAATGCCAAATGGTTCGCCAGGATTTGATTTCCACC 900
Db 895 TACAACCTACATGAAGTTTGCCAAACCAATGCCAAATGGTTCGCCAGGATTTGATTTCCACC 954
QY 901 TGCAACAGACAACCCGCACTAGCTGACTACGGCTCTCGCGGAGAGCCCAAC 960
Db 955 TGCAACAGACAACCCGCACTAGCTGACTACGGCTCTCGCGGAGAGCCCAAC 1014
QY 961 ATGTGCAGGACAAATGTTAGGGGCCATACCTACGCCCTTTGCTGGTGGTGTGTATGAT 1020
Db 1015 ATGTGCAGGACAAATGTTAGGGGCCATACCTACGCCCTTTGCTGGTGGTGTGTATGAT 1074
QY 1021 ATTGGGCATCCATATGATGACCCGACTCCGCCAAAGTTATTACAAATAATTTCTGGCAAAG 1080
Db 1075 ATTGGGCATCCATATGATGACCCGACTCCGCCAAAGTTATTACAAATAATTTCTGGCAAAG 1134
QY 1081 GACTCTGTGATGAGCTATCGGGGTCAACATCAACTACACCAGTCCCAATTAATGACGTC 1140
Db 1135 GACTCTGTGATGAGCTATCGGGGTCAACATCAACTACACCAGTCCCAATTAATGACGTC 1194
QY 1141 TACTACGCTTCCAGCAAAAGGCGACTTGTCTGGGCCAACTTTCATCGAAGACTCGAG 1200
Db 1195 TACTACGCTTCCAGCAAAAGGCGACTTGTCTGGGCCAACTTTCATCGAAGACTCGAG 1254
QY 1201 GAGATCCTTGTCTCCCGGTGCGTGTCTCCCTCACTATGCGGAGCGCGATTACATCTGC 1260
Db 1255 GAGATCCTTGTCTCCCGGTGCGTGTCTCCCTCACTATGCGGAGCGCGATTACATCTGC 1314
QY 1261 AACTGGTTCGCGGTGAGCGGTTTCCTCGCTCGCAACTACTCCGAGCGCGCCAGTTC 1320

Db 1315 AACTGGTTCGCGGTGAGCGGTTTCCTCGCTCGCAACTACTCCCAAGCGCGCCAGTTC 1374
QY 1321 CGAAGCGCAGGTTACACGCCCTGAAAGTCAACGCGCTCGAGTATGGGCAAACTCGCGAG 1380
Db 1375 CGAAGCGCAGGTTACACGCCCTGAAAGTCAACGCGCTCGAGTATGGGCAAACTCGCGAG 1434
QY 1381 TATGGTAATTTCTCCTTCACTCGCGTCTATGAGSCAGGCCATGAAGTCCCATACTACCAG 1440
Db 1435 TATGGTAATTTCTCCTTCACTCGCGTCTATGAGSCAGGCCATGAAGTCCCATACTACCAG 1494
QY 1441 CCCATCGGCTCCCTGCAATTTTAAACCGGACTATCTTCGGTGGGATATGCGAGAGGC 1500
Db 1495 CCCATCGGCTCCCTGCAATTTTAAACCGGACTATCTTCGGTGGGATATGCGAGAGGC 1554
QY 1501 CAGAAGAGATCTGSCCCAGCTACAAGACGAAATGGAACGGCTACAGCTACGCATACACAG 1560
Db 1555 CAGAAGAGATCTGSCCCAGCTACAAGACGAAATGGAACGGCTACAGCTACGCATACACAG 1614
QY 1561 TCGTCCGTGCGCTGCGCTACGGCTACCGGCTACCGGCTACGCTGTGTGATG 1608
Db 1615 TCGTCCGTGCGCTGCGCTACGGCTACCGGCTACCGGCTACGCTGTGTGATG 1662

Search completed: September 16, 2003, 19:15:52
Job time : 554 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 14:42:53 ; Search time 554 Seconds
(without alignments)
7835.192 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 1608
Sequence: 1 cttccaggagatcacccgc.....gcattgtccaggttggtatg 1608

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	1608	100.0	1662	19 AAV28620	A. oryzae ATCC2038
2	874.8	54.4	1656	24 ABZ78288	A. niger serine ca
3	313.6	19.5	3150	24 ABZ78231	A. niger serine ca
4	164.8	10.2	1007	25 ABZ51997	Aspergillus oryzae
5	134.4	8.4	1872	24 ABZ78289	A. niger serine ca
6	118.6	7.4	1665	24 ABZ78283	A. niger serine ca
7	116.8	7.3	1581	24 ABZ78241	A. niger serine ca
8	116.8	7.3	2940	24 ABZ78184	A. niger serine ca

9	114.6	7.1	3080	24 ABZ78226	A. niger serine ca
10	107.8	6.7	3221	24 ABZ78232	A. niger serine ca
11	103	6.4	678	21 AAF12116	Aspergillus oryzae
12	87.8	5.5	636	21 AAF13804	Aspergillus oryzae
13	86.4	5.4	2503	15 AAQ55347	Sequence of gene K
14	84	5.2	2002	17 AAT28284	A. niger SFAG 2 ca
15	82.4	5.1	1611	24 ABZ78243	A. niger carboxype
16	78.2	4.9	2068	17 AAT28283	A. niger carboxype
17	78.2	4.9	2660	24 ABZ78186	A. niger Bo-1 geno
18	75.6	4.7	626	21 AAF12522	A. niger carboxype
19	68.4	4.3	1368	24 ABZ78267	Aspergillus oryzae
20	68.4	4.3	1653	24 ABQ76547	C. albicans BAX-as
21	68.4	4.3	2441	24 ABZ78210	A. niger carboxype
22	63.2	3.9	1446	24 ABZ78269	A. niger carboxype
23	61.8	3.8	2027	24 ABQ76315	S. cerevisiae BAX-
24	61.8	3.8	1742	23 ABL07915	Drosophila melanog
25	58.6	3.6	3826	23 ABL07914	Drosophila melanog
26	58.6	3.6	4509	23 ABL08118	Drosophila melanog
27	57.4	3.6	475	25 ABZ54889	Aspergillus oryzae
28	56.4	3.5	1814	22 AAF68401	Human lung tumor
29	56.4	3.5	1814	24 ABK38312	cDNA encoding clon
30	56.4	3.5	1814	25 ACA10641	Human lung cancer-
31	56.4	3.5	1814	25 ABX99592	Lung cancer therap
32	56.4	3.5	1815	24 ABQ88162	Human osteoblast d
33	56.4	3.5	1815	24 ABN95798	Gene #2296 used to
34	56.4	3.5	1815	24 ABN95798	Lung cancer associ
35	54.8	3.4	2864	21 AAF18312	Human dithp enzyme
36	49.2	3.1	1997	25 ACC46113	A. niger carboxype
37	48.4	3.0	1839	24 ABZ78258	A. niger carboxype
38	48.4	3.0	2730	24 ABZ78201	Trichoderma reesei
39	48	3.0	869	21 AAF14932	Randomising oligon
40	47.6	3.0	390	13 AAQ21833	PCR primer for 5'
41	47.6	3.0	390	14 AAQ36859	Sequence containin
42	47.6	3.0	390	22 AAF76910	Oligo #7 for cloni
43	47.6	3.0	390	24 AAI72775	Human hydrophobic
44	47.6	3.0	1428	22 AAF94477	Human serine carbo
45	47.6	3.0	1551	19 AAV64076	

ALIGNMENTS

RESULT 1
AAV28620
ID AAV28620 standard; DNA; 1662 BP.
XX
AC AAV28620;
XX
DT 27-AUG-1998 (first entry)
XX
DE A. oryzae ATCC20386 carboxypeptidase I DNA.
XX
KW Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous;
KW food industry; ss.
XX
OS Aspergillus oryzae.

Key Location/Qualifiers
CDS 1..1662
FT /*tag= a
FT /product= carboxypeptidase I
FT /note= "partial coding sequence"

WO9814599-A1

09-APR-1998

03-OCT-1997; 97WO-US17977.

27-NOV-1996; 96US-0757534.

04-OCT-1996; 96US-0726880.

(NOVO) NOVO NORDISK BIOTECH INC.

10262, 100%

PA (NOVO) NOVO-NORDISK AS.
 XX Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;
 PI Klotz A, Mathisen TE, Rey M;
 XX WPI; 1998-240098/21.
 DR P-PSDB; AAW56099.
 XX Carboxypeptidase from *Aspergillus oryzae* - hydrolyses proteinaceous
 PT substrates, useful for improving flavour of foods
 XX Claim 2; Fig 3; 82pp; English.
 XX This DNA sequence encodes carboxypeptidase I from *Aspergillus oryzae*.
 CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
 CC 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a
 CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and
 CC 60 deg. C. It also has the capacity to hydrolyse X from N-Cbz-Ala-X where
 CC N-Cbz is N-carboxybenzoyl and X is any amino acid. The carboxypeptidases
 CC can be used for obtaining hydrolysates (which can be enriched in free
 CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous
 CC substrates. The carboxypeptidases can be used in flavour-improving
 CC compositions in the food industry. The products can also be used for the
 CC production of polypeptides free of carboxypeptidase activity.
 XX Sequence 1662 BP; 396 A; 468 C; 398 G; 400 T; 0 other;
 SQ

Query Match 100.0%; Score 1608; DB 19; Length 1662;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCCAGGAAGTACACCGGCTCGTGGTAGAGACAGTACCCAGAAACCCACCGGG 60
 DB 55 CTTCCAGGAAGTACACCGGCTCGTGGTAGAGACAGTACCCAGAAACCCACCGGG 114
 QY 61 GTCAGAGCTCTTAACACCGGAAACAAATGTCACCATCCGGTACAGGAACCCGGGACAG 120
 DB 115 GTCAGAGCTCTTAACACCGGAAACAAATGTCACCATCCGGTACAGGAACCCGGGACAG 174
 QY 121 GGGCTGTCGAGACTACCCGGGTCAAAATCTTCTGATATGTCGACACTCTCC 180
 DB 175 GGGCTGTCGAGACTACCCGGGTCAAAATCTTCTGATATGTCGACACTCTCC 234
 QY 181 GAGTCCCATACCTTCTTCTGTTCTTGAAGCCAGACATAACCCAGAACTGCACCTATC 240
 DB 235 GAGTCCCATACCTTCTTCTGTTCTTGAAGCCAGACATAACCCAGAACTGCACCTATC 294
 QY 241 ACATTGGTGTGAATGGTGGCCCTGGAAGCCGATTTTGTATCGGTCTCTTGAAGAGTTG 300
 DB 295 ACATTGGTGTGAATGGTGGCCCTGGAAGCCGATTTTGTATCGGTCTCTTGAAGAGTTG 354
 QY 301 GGGCTTCCCATGTCAAATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGAAACGAG 360
 DB 355 GGGCTTCCCATGTCAAATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGAAACGAG 414
 QY 361 GTCTCCAAATTAATTCCTTCCAGCCATTTGGAGTGGGCTTTTCATATAGTATAG 420
 DB 415 GTCTCCAAATTAATTCCTTCCAGCCATTTGGAGTGGGCTTTTCATATAGTATAG 474
 QY 421 GTTGTATGGTTCATTAACCTTGTAATCGGGTCTCGAAATTCGAGCTTTTCAGAGTT 480
 DB 475 GTTGTATGGTTCATTAACCTTGTAATCGGGTCTCGAAATTCGAGCTTTTCAGAGTT 534
 QY 481 CAGGCGCGGTACCCAAACCATGATGCCACTCTGATCGATACTACCAATCTTCCCGCAGAG 540
 DB 535 CAGGCGCGGTACCCAAACCATGATGCCACTCTGATCGATACTACCAATCTTCCCGCAGAG 594
 QY 541 GCCGCTTGGGAGATCTCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTG 600
 DB 595 GCCGCTTGGGAGATCTCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTG 654
 QY 601 CAGTCTAAGGACTTCAGTCTATGGAGCGAGACTATGGAGGSCACTATGGTCTCTGCATTC 660
 XX

RESULT 2

ABZ78288

ID ABZ78288 standard; cdna; 1656 BP.

XX

AC ABZ78288;

XX

DB 655 CAGTCTAAGGACTTCAGTCTATGACGAGAGCTATGAGGAGGCACTATGGTCTCCGCAATC 714
 QY 661 TTCAATCATTTTTTACGAGCAGAAATGCCAAGATTCGCAACGCTAGCTGTATTAATGGTTCAG 720
 DB 715 TTCAATCATTTTTTACGAGCAGAAATGCCAAGATTCGCAACGCTAGCTGTATTAATGGTTCAG 774
 QY 721 CTTAATTTCAACTCTCTGGGAAATTTAATACGGCATCAPCGACGAGGCGATCCAGCCGCT 780
 DB 775 CTTAATTTCAACTCTCTGGGAAATTTAATACGGCATCAPCGACGAGGCGATCCAGCCGCT 834
 QY 781 TACTACCTGTAATTCGCTGTCAACAATACCTTCAAGGTATCAAGGCTGTCAACGACCGCTC 840
 DB 835 TACTACCTGTAATTCGCTGTCAACAATACCTTCAAGGTATCAAGGCTGTCAACGACCGCTC 894
 QY 841 TACAACATACATGAAGTTTGCAACCAATGATGCAATGTTGCCAGGATTTGATTTCCACC 900
 DB 895 TACAACATACATGAAGTTTGCAACCAATGATGCAATGTTGCCAGGATTTGATTTCCACC 954
 QY 901 TGCAGACAGACAAACCCGACCGCATAGCTGACTACGCCCTCTGGCGGAGGACCAAC 960
 DB 955 TGCAGACAGACAAACCCGACCGCATAGCTGACTACGCCCTCTGGCGGAGGACCAAC 1014
 QY 961 ATGTGACGAGGACAATGTTGAGGGGCCATACCTACGCTTGTGCTGCTGGTGTGTATGAT 1020
 DB 1015 ATGTGACGAGGACAATGTTGAGGGGCCATACCTACGCTTGTGCTGCTGGTGTGTATGAT 1074
 QY 1021 ATTCGGCATCATATGATGATGACCCGACTCCGCCAAGTTATTACAACAATTTCTGGCAAG 1080
 DB 1075 ATTCGGCATCATATGATGATGACCCGACTCCGCCAAGTTATTACAACAATTTCTGGCAAG 1134
 QY 1081 GACTCTGTATGAGGCTATCGGCTCAACATCACTACCCAGTCCCAATTAATGACCTC 1140
 DB 1135 GACTCTGTATGAGGCTATCGGCTCAACATCACTACCCAGTCCCAATTAATGACCTC 1194
 QY 1141 TACTACGCTTTCACGACAAACAGGCGACTTTGCTGGCCCACTTCATCGAAGACCTCGAG 1200
 DB 1195 TACTACGCTTTCACGACAAACAGGCGACTTTGCTGGCCCACTTCATCGAAGACCTCGAG 1254
 QY 1201 GAGATCCCTGCTCTCCCGTGGTGTCTCCCTCATCTATGGCGAGCGGATTTACATCTGC 1260
 DB 1255 GAGATCCCTGCTCTCCCGTGGTGTCTCCCTCATCTATGGCGAGCGGATTTACATCTGC 1314
 QY 1261 AACTGGTTCGGGCTCAGGCGTTTCCCTCGCTGCGAACTACTCCAGGCGCCCGAGTTC 1320
 DB 1315 AACTGGTTCGGGCTCAGGCGTTTCCCTCGCTGCGAACTACTCCAGGCGCCCGAGTTC 1374
 QY 1321 CGAAGCGAGGCTACACGCCCTGAAAGTCAACGGGCTCGAGTATGGGAAACTCGCGAG 1380
 DB 1375 CGAAGCGAGGCTACACGCCCTGAAAGTCAACGGGCTCGAGTATGGGAAACTCGCGAG 1434
 QY 1381 TATGTAATTTCTCTCACTCGGCTATGAGGAGGCGCATGAGTCCCATACTACAG 1440
 DB 1435 TATGTAATTTCTCTCACTCGGCTATGAGGAGGCGCATGAGTCCCATACTACAG 1494
 QY 1441 CCCATCGCTCCCTGCAATGTTTAAACGGGCTATCTTCGGTTGGATATCGCAGAGGC 1500
 DB 1495 CCCATCGCTCCCTGCAATGTTTAAACGGGCTATCTTCGGTTGGATATCGCAGAGGC 1554
 QY 1501 CAGAAGAAGATCTGGCCAGCTACAGAGCAATGGAACGGCTACAGCTACGATACAG 1560
 DB 1555 CAGAAGAAGATCTGGCCAGCTACAGAGCAATGGAACGGCTACAGCTACGATACAG 1614
 QY 1561 TCGTCCGTGCGCTGCTACCGCTACCGCTACCGCATGTCCAGTGTGGTATG 1608
 DB 1615 TCGTCCGTGCGCTGCTACCGCTACCGCTACCGCATGTCCAGTGTGGTATG 1662

Claim 1; Page 187-189; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A polypeptide or polynucleotide of the invention is useful for diagnosing a fungal infection such as aspergillosis, or as a query sequence to perform a search against public databases. A polypeptide of the invention is useful in a selected number of industrial or pharmaceutical processes, in laboratory or clinical processes, in food industry (baking, brewing, cheese manufacturing, meat tenderising), in tanning industry and in the manufacture of biological detergents. A polypeptide may also be useful for improving protein solubility, extraction yields, viscosity or taste, texture, nutritional value, minimising of antigenicity or anti-nutritional factors, colour or functionality as well as processing aspects like filterability of the proteinaceous raw material. The sequences shown in AB278180-AB278236 represent genes encoding the A. niger proteases of the invention.

Sequence 3150 BP; 801 A; 778 C; 738 G; 833 T; 0 other.

Query Match	19.5%;	Score 313.6;	DB 24;	Length 3150;
Best Local Similarity	72.5%;	Pred. No. 4.2e-83;		
Matches 406;	Conservative	0;	Mismatches 154;	Indels 0;
QY	1037	ATCACCCGGCTCCGCCAAGTATTATACACAAATTTCTGGCAAGGACTCTGTCATNGAAG	1096	
Db	2194	AGGACCCGGACCCCGCGCTCTACTTTGTTGCATACCTCAAGAAGACTCACTCATGGATG	2253	
QY	1097	CTATCGCGGTCAACATCAACTACACCCAGTCCAAATATGATGCTACTACGCTTCTCCAGC	1156	
Db	2254	CTATCGCGGTGGACATTAACTACACCGAGTCCAGCGCGGAAGTATATTATGCAATCCAGC	2313	
QY	1157	AAACAGCGGACTTTGCTGGCCCACTTCACGAAAGACCTCGAGGAGATCCTTCTCTCC	1216	
Db	2314	AGACCGGCGACTTTGATGSCCGAAITTCATTAGGACCTCGAAGAGATCCTCCAACTCC	2373	
QY	1217	CCGTGCGTGTCTCCCTCATCTATGCGGACGCCGATACATCTCAACTGCTGTTCCGCGGTC	1276	
Db	2374	CGGTACGCTGTCTGTGATCTACGGCGATGCCGACTAATCTCTAACTGGTTCGGCGGTC	2433	
QY	1277	AGGCCGTTTCCCPGCTCGGAACACTCTCCAAAGCCGCCAGTTCCGAAGCGCAGGGTACA	1336	
Db	2434	AGGCCATCTCACTCGCAGTTAACTACGCCCATGCAGCTCAGTTCGCTGCAGCGGATACA	2493	
QY	1337	CGCCCTCGAAGTCAACGGCGTCCAGTATGGGAAACTCCGAGTAGTGTAAITTCCT	1396	
Db	2494	CACCCATGACAGTAGATGGGTTCGAATACGGTGAGACTCGCGAGTAGGCAACTTTTCGT	2553	
QY	1397	TCATCGCGCTCTATGAGCGAGGCCATGAAGTCCCATCTACCAAGCCCATCGCTCCCTGC	1456	
Db	2554	TCACCCGCGTATATCAGGCTGGSCACAGGTTCCATCTATCAACCGATCGCAGCTTGC	2613	
QY	1457	AAATGTTTAAACGGACTATCTTCGGTTGGGATATCCGAGGGCCAGAAGAGACTTCGC	1516	
Db	2614	AGCTGTTTCAACCGTACTTTATTGGATGGGATATTCAGCGGGTACAAGTTCAGTTGGC	2673	
QY	1517	CCAGGTACAAAGCGAATGAACGGGTACAGTACGCATACACAGTCTCGTCCGTCGCCGTC	1576	
Db	2674	CCGAATATAGCAACACGGGACATCCAGGCTACACACGAGTCTGTCGTGCCACTGT	2733	
QY	1577	CTACGGCTACCGAGCATGTCC	1596	
Db	2734	CCACGGCGTCGAGTACCAC	2753	

RESULT 4	
ABZ51997	
ID	ABZ51997 standard; cDNA; 1007 bp.
XX	
XX	
AC	ABZ51997;
XX	
XX	
DT	28-MAR-2003 (first entry)
XX	

DE Aspergillus oryzae polynucleotide SEQ ID NO 1110.

XX

KW Aspergillus oryzae; fermentation; fungus; industrial; EST;

KW expressed sequence tag; gene; ss.

XX

OS Aspergillus oryzae.

XX

PN WO200279476-A1.

XX

PD 10-OCT-2002.

XX

PF 22-MAR-2002; 2002WO-IB00890.

XX

PR 30-MAR-2001; 2001JP-0098371.

XX

PA (NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

PA (NARE-) NAT RES INST BREWING.

PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX

PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;

PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX

DR WPT; 2003-045817/04.

XX

PT Detection of expression of specific Aspergillus genes for monitoring

PT the fermentation and growth conditions of the fungus, using DNA probes

XX

PS Claim 1; SEQ ID NO 1110; 48pp + Sequence Listing; Japanese.

XX

CC The invention relates to a polynucleotide having any of 6006 specific

CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under

CC specific culture conditions including one or more of eutrophic,

CC oligotrophic, solid, early germination, alkaline, high temperature, low

CC temperature or maltose culture or polynucleotides stringently hybridising

CC to these sequences. The polynucleotides are useful for monitoring the

CC progress of fermentation and the growth conditions of a fungus,

CC especially of Aspergillus oryzae which is widely used in industrial

CC fermentation. Also monitoring for fungal contamination.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

```

SQ Sequence 1007 BP; 253 A; 270 C; 257 G; 224 T; 3 other;

Query Match      10.2%; Score 164.8; DB 25; Length 1007;
Best Local Similarity 52.4%; Pred. No. 1.1e-38;
Matches 481; Conservative 3; Mismatches 390; Indels 44; Gaps 4;

QY 630 GAGCTATGGAGGGCAGCATATGTCGICGATCTCTTCAATCATTTTTACGACGAGAAGAGAG 689
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 42 GAGCIAGGAGAGGGTACGGCCATCTTTCAGTGCITTCCTCCAGGAGCAGAAAGAA 101

QY 690 AATTGCCAACGGTAGTGTTA-----ATGGTGTCAGCTTAATTTCAACTCTCTGGGA 741
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 102 GATTGCCAACGSTCCAATTGATCATCGAGATGCTCACTATAITCACTGGACACCTTCGGA 161

QY 742 ATTATTAACGGATCATCTPCGACGAGGGGATCCAGGCGCCCTTACTACCTGAATTCGTGTG 801
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162 ATTATCAATGGCTCGCTCGATCTCTGGTCACTCGCCATCAATATCCCGAGATGCATAC 221

QY 802 ARCAAAATACCTACGGTATCAGGCGTCACGACGAGCGGTCTACAACATACATGAGTTTGGC 861
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222 AACAAATACCTATGGCAITCAGGCGTATCAACAAACTGTATATGATGCGGATGAGGCT 281

QY 862 AACCAANTGCCAATAGTTGCCAGGATTTGATTTCACCTG----- 902
      || || || || || || || || || || || || || || || || || || || || ||
Db 282 TGGAGCAAGCCGAGGAGGCTGCAAGSATCATGATCGAATGTGTCGGTTGGCTGCAGAG 341

QY 903 -----CAAAACAGACAACCGCACCGCATTTAGCTGACTACGCCCTCTCGGCCGAGCCACC 957
      || || || || || || || || || || || || || || || || || || || || ||
Db 342 GGAGACCCCGAAATGATPACGGCACAATAGACAGCGTCAACAAGGTCTGTGCGAAGCAAC 401

```

QY	958	ARCATGTGAGGAGCAAAATGTTGAGGGCCATACATAGCCCTTGTGCTGGTGTGTAT	1017	PR	26-FEB-2001; 2001EP-0200708.
Db	402	AACTACTGAGCAACCAAGTAGAAGGCCCATATCTACTCTGACGAGGCTATTAC	461	PR	26-FEB-2001; 2001EP-0200719.
QY	1018	GATATTCGCATCCATATGATGACCGGACTCCGCAAGTATTACAAATAATTCGGCA	1077	PR	28-MAR-2001; 2001EP-0000075.
Db	462	GACATCTGCATTTTCGACCCGATCCTTTCCGCCCTCCGTACTACTTTGGTTCCTTAAAC	521	PR	28-MAR-2001; 2001EP-0000078.
QY	1078	AAGGACTCTGTGATGAGACCTATCGGCGTCAACATCACTACACCCAGTCCATAATGAC	1137	PR	28-MAR-2001; 2001EP-0000080.
Db	522	CAGCACTGGGTCCAGAGGCCCTCGGGGTTCGGGTAAATTCACCGAATCTGTGATGAGC	581	PR	28-MAR-2001; 2001EP-0000087.
QY	1138	GTCTACTAGCTTTCCAGCAAAACAGCGGACTTTTGTGCGCCCAACTTCATCGAAGACCTC	1197	PR	28-MAR-2001; 2001EP-0000088.
Db	582	GTCTACAATGGCTTTTCCGCAACAGGTGACTACCCACGCTCCGATGACGCGGTACCTG	641	PR	21-MAY-2001; 2001EP-0000156.
QY	1198	GAGGAGATC-----CTTGTCTCCCGTGCAGGCGGTTTCCCTGCTGCGAACTACTCCCA	1248	PR	21-MAY-2001; 2001EP-0000159.
Db	642	GAGGATATGCATACGCTGCTTACTCTGGCATCAAAATTCGCCCTAGTATGCGGATCGG	701	PR	21-MAY-2001; 2001EP-0000160.
QY	1249	GATTACATCTGCAACTGGTTTCGGCGTACAGGCGGTTTCCCTGCTGCGAACTACTCCCA	1308	PR	21-MAY-2001; 2001EP-0000162.
Db	702	GATTACGATGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	761	PR	21-MAY-2001; 2001EP-0000165.
QY	1309	GCGGCCAGTTCGGAAGCGGAGGATACAGCCCTGAAAGTAAACGCG---GTCGAGTAT	1365	PR	21-MAY-2001; 2001EP-0000166.
Db	762	GCGGCCAAGTTCGCGCTCGGTTTACGCCCTCTGAAGACCAATGCTCTATATGATAGT	821	PR	21-MAY-2001; 2001EP-0000168.
QY	1366	GGGGAACCTCGGAGATGATGTAATTTCTTCCTTCACGCGGTCTATGAGGAGGCGCATGAA	1425	PR	21-JUN-2001; 2001EP-0000240.
Db	822	GGCTGGTGCAGAGTGCAGGAAATCTCTGCTCACTCGGCTTTTCGAAAGCGGTCTATGAG	881	PR	21-JUN-2001; 2001EP-0000242.
QY	1426	GTCCCATACTACAGCCCATCGCTCCCTGCAATTTGTTAAACCGACTATCTTCGGTTGG	1485	PR	21-JUN-2001; 2001EP-0000244.
Db	882	GTCCGGCATATCAGCTGCTGAACGGCGATGAGATCTCCACCGAGCATTTGTTCAACAGA	941	PR	21-JUN-2001; 2001EP-0000246.
QY	1486	GATATCGCAGAGGCCAG 1503		PR	12-JUL-2001; 2001EP-0000280.
Db	942	GACATTCGAGCGGCAAG 959		PR	30-JUL-2001; 2001EP-0000323.
RESULT 5				PR	02-AUG-2001; 2001EP-0000341.
ABZ78289				PR	02-AUG-2001; 2001EP-0000342.
AC				PR	02-AUG-2001; 2001EP-0000343.
AC				PR	02-AUG-2001; 2001EP-0000344.
AC				PR	09-AUG-2001; 2001EP-0000357.
XX				PR	16-AUG-2001; 2001EP-0000374.
DT				PR	16-AUG-2001; 2001EP-0000377.
XX				PR	20-SEP-2001; 2001EP-0000478.
DE				PR	20-SEP-2001; 2001EP-0000483.
XX				PR	22-OCT-2001; 2001EP-0000552.
KW				PR	22-OCT-2001; 2001EP-0000553.
KW				PR	22-OCT-2001; 2001EP-0000554.
KW				PR	22-OCT-2001; 2001EP-0000556.
XX				PR	22-OCT-2001; 2001EP-0000557.
OS				PR	22-OCT-2001; 2001EP-0000558.
XX				PR	15-NOV-2001; 2001EP-0000454.
XX				PR	21-DEC-2001; 2001EP-0005117.
FT				PA	(STAM) DSM NV.
FT				PI	Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
FT				PI	Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
FT				PI	Maier D, Spreafico F, Folkers J, Hopper S, Kemmer W, Tan P;
FT				PI	Stiebler J, Albang R;
XX				XX	WPI: 2002-723203/78.
XX				DR	P-PSDE: ABR38865.
XX				DR	Novel isolated protease polypeptide useful in laboratory, clinical,
XX				DR	pharmaceutical, chemical, diagnostic, personal care and industrial
XX				DR	applications
XX				XX	Claim 1; Page 253-254; 394pp; English.
XX				XX	The invention relates to a novel isolated protease polypeptide. A
XX				XX	polypeptide or polynucleotide of the invention is useful for diagnosing a
XX				XX	fungal infection such as aspergillosis, or as a query sequence to perform
XX				XX	a search against public databases. A polypeptide of the invention is
XX				XX	useful in a selected number of industrial or pharmaceutical processes, in
XX				XX	laboratory or clinical processes, in food industry (baking, brewing, in
XX				XX	cheese manufacture, meat tenderising), in tanning industry and in the
XX				XX	manufacture of biological detergents. A polypeptide may also be useful
XX				XX	for improving protein solubility, extraction yields, viscosity or taste,
XX				XX	texture, nutritional value, minimising of antigenicity or
XX				XX	anti-nutritional factors, colour or functionality as well as processing
XX				XX	aspects like filterability of the proteinaceous raw material. The
XX				XX	sequences shown in ABZ78237-ABZ78293 represent cDNA encoding the
XX				XX	A. niger proteases of the invention.
XX				XX	Sequence 1872 BP; 402 A; 434 C; 570 G; 466 T; 0 other;

Query Match	8.4%	Score 134.4	DB 24	Length 1872
Best local Similarity	48.1%	Pred. No. 2e-29		
Matches 644	Conservative 0	Mismatches 661	Indels 33	Gaps 8
QY	186	CCATACCTCTCTCTGGTCTTCGAGACCCAGACAT-----AACCCAGAAACGACACCAT	239	
DB	282	CATYACCTTTTCTGGTACTTTCTTCGGCGCATCACCAACATGATATACATCCCACT	341	
QY	240	CACATPTGGTGTGAATGGTGCCCTGGAAGCGATCTTTTGATCGTCTCTTCGGAAGGT	299	
DB	342	CACCATCTGGATGAACGCGCGCGGGATCTCTCAATGATGGGCHATTTCAGAGAA	401	
QY	300	GGGCCCTTGCCATGTCAATTCGACTTTTGAATGACTACATCAACCTCAC---TCGTGAA	356	
DB	402	CGGCCATGTACTGTGAATAGCGACTCGAATTCACGGCTTAATATCCCTGGTCTGGAA	461	
QY	357	CGAGTCTCCAAATTTACTATTCTCTGCCAGCCATTGGAGTCGCTTTTCATATAGTGA	416	
DB	462	TGAGTACGTTCGATATGTTGATATTGACAGCGCGGTGCAGACGGGATTTAGTTATGAT	521	
QY	417	TACGGTTGATGGGTCCATTAACCCCTGTAACCTGGGTCTGTGGAATTCGACCTTGCAGG	476	
DB	522	GTTGAGGAAAGGACGCTAGAT---TTGAATGAGACGTTTGGTGGGACGTTCGCG-576		
QY	477	AGTTTCAGGCGCGTACCCAAACCATTGATGCCACTCTGATCGATCTACCAATCTTTCGCCG	536	
DB	577	-AGTCAGATGTGCAUGGACGGGTGAATGGGACGGTTAATGGGGAAGGCGCTTTGGGT	635	
QY	537	AGAGCCCGCTTGGGAGATCTCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACICTAG	596	
DB	636	TGCGTTGCAGGTTGGTGGTGAACTCTCTGGAATGTTTCTCTGTGTGACGGGAATGG	695	
QY	597	GGTCAGCTTAAGGACTTCAGTCTATGACGGAGAGCTATGAGGGGCACTATGTCCTGC	656	
DB	696	TGCTGTGTGATCACAGGTGAGTATATGACGGAGTCAATGAGGACGTAIGGACCGCG	755	
QY	657	ATTCCTTCAATCAITTTTACGAGCAGATGAGAGAAATTCGCAACGGTAGTGT---TAATGG	713	
DB	756	ATACACGGCGCTTCTCAGAGATGAATGAGAGGATTCAGAGTGGGAGGTAAAGCACCGG	815	
QY	714	TGTTTCAGCTTAATTTCAACTCTCTGGGAATTTATTAACGGCATCATGACGAGGCGATCCA	773	
DB	816	GAAGAAGATCCATTTGGATACGCTGGGCAATTTCAATGGTGTGTGATTTACTCTGTGCA	875	
QY	774	GGCCCTTACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGCGTGTCAACGA	833	
DB	876	GGTCCCTTCGCTCCCTGAGCAGCGGTATAACAATACGATGGGATCGAGGGAATCAATCG	935	
QY	834	GACCGTCTCAACTACATGAATGATTTCCCAACCAAAATGCCAAATGGTTGTCAGGATTTGAT	893	
DB	936	CACGCTCTAGCACGGGCTATGGATAGTTGGAGCAAGCCTGGGCGGTGTCAAGGAATGAT	995	
QY	894	TTCCACTCGAAACAGACAAACGCGCATTTAGCTGACTACGCCCTCTGCGCCGAGCG	953	
DB	996	CATCGAGTGTCCGGATGCTGGCGAGCTGG---AGATCCCTCATCATCTCCGAGGAGCG	1052	
QY	954	CACCACATGTGAGGAGCAATGTTGAGGGGCCATACTAGCCCTTTTCTGCTGTGCTGTGT	1013	
DB	1053	GTCCGACTACTGTTCCGGGAGATCAAGACCCGTATACGAATACCTCCGGCGGAGATA	1112	
QY	1014	GATGATATTGGCATCCCATATGATCACCCGACTCCGCAAGTANTTACAAAAATTTCT	1073	
DB	1113	CTAGCATATAGCGCATTTACGCCCGGATGACGCTCTCGTGCCTTACTTCTCGCGGTTCTT	1172	
QY	1074	GGCAAAGGACTCTGTTCATGACGCTATCGCGGTCAACATCACTACACCGACGATCCATAA	1133	
DB	1173	GAATCGCCCATGGTCCAAAGGCATTTGGGHTCCCGGTGAACATATACCATGTCGTCAGA	1232	
QY	1134	TGACGTCTACTAGCTTTCCAGCAAAACAGCGACTTTGTCTGSCCAACTTCAATCAAGA	1193	
DB	1233	GGCAATPGGGGAACAGTTTCGCTCGAGCGGCGATTATCCGGAATATGATCCCGCGGAAT	1292	
QY	1194	CCTCGAGAGA-----TCCITGCTCTCCCGTGGGTGCTCCCTCATCTATGAGGGA	1244	

Db	1293	GATCGGGATATIGGATACTTGGTATGCTTGACTCCGGGTCAAGGTGGCIATGGTATATGGGA	1352
QY	1245	CGCGSANTACATCTGCAACTGGTTCGGGGTCAGGCGGTTTCCTCGCTGCGGAACIACTC	1304
Db	1353	CCGGGACTATGCTTTCGCTGCGGGCGGGGAAGATGTACGCTGCTGGTGGAGTACGA	1412
QY	1305	CCAGCCGCCACAGTTCGGAAGCGCAGGAGTACACCCCTGAAAGTCAACGCGCTCGAGTA	1364
Db	1413	GGATCCGGAGAGTTCGCTGCTGGGTATGCCGAAGTGCAGACGAAGTCATCCTACGT	1472
QY	1365	TGGGG--AACTCGCAGTATGTAATTTCTCTTCACTCGGCTCTATCAGGCAGGCCA	1421
Db	1473	TGGGGGCTCTAGTAGGAGTATGGGAATTTCTGCTTCAACGCGTCTTTCAGCGGGCCA	1532
QY	1422	TGAAGTCCCACTACCAAGCCATCGCTCCCTCAATTTTAAACGGACTATCTTCGG	1481
Db	1533	TGAGTGCCATTTATCAGCCGAACGGCGTATGAGATTTTAAATCGCTCAGTTAA	1592
QY	1482	TTGGGATATCGCAGGGG	1499
Db	1593	TTGGGATATCGCAGGGG	1610
RESULT 6			
ABZ78283			
ID	ABZ78283 standard; cDNA; 1665 BP.		
XX	AC ABZ78283;		
XX	24-APR-2003 (first entry)		
DE	A. niger serine carboxypeptidase cDNA #2.		
KW	Protease; fungal infection; aspergillosis; food; tanning; detergent;		
KW	protein solubility; viscosity; taste; texture; nutritional value;		
XX	gene; ss.		
OS	Aspergillus niger.		
Key	Location/Qualifiers		
CDS	1..1665		
FT	/*tag= a		
FT	/EC_number= "3.4.16.6"		
XX	WC200268623-A2.		
PD	06-SEP-2002.		
XX	22-FEB-2002; 2002WO-EP01984.		
XX	23-FEB-2001; 2001EP-0200657.		
PR	23-FEB-2001; 2001EP-0200658.		
PR	23-FEB-2001; 2001EP-0200660.		
PR	26-FEB-2001; 2001EP-0200706.		
PR	26-FEB-2001; 2001EP-0200707.		
PR	26-FEB-2001; 2001EP-0200708.		
PR	26-FEB-2001; 2001EP-0200719.		
PR	28-MAR-2001; 2001EP-0000075.		
PR	28-MAR-2001; 2001EP-0000078.		
PR	28-MAR-2001; 2001EP-0000080.		
PR	28-MAR-2001; 2001EP-0000087.		
PR	28-MAR-2001; 2001EP-0000088.		
PR	21-MAY-2001; 2001EP-0000156.		
PR	21-MAY-2001; 2001EP-0000159.		
PR	21-MAY-2001; 2001EP-0000160.		
PR	21-MAY-2001; 2001EP-0000162.		
PR	21-MAY-2001; 2001EP-0000165.		
PR	21-MAY-2001; 2001EP-0000166.		
PR	21-MAY-2001; 2001EP-0000168.		
PR	21-JUN-2001; 2001EP-0000240.		
PR	21-JUN-2001; 2001EP-0000242.		
PR	21-JUN-2001; 2001EP-0000244.		

PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 30-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 02-AUG-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 15-NOV-2001; 2001EP-0004464.
PR 21-DEC-2001; 2001EP-0005117.
XX (STAM) DSM NV.
XX PA
XX PI Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albarg R;
XX PI
DR WPI: 2002-723203/78.
DR P PSDB; ABR38859.
XX PT Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications
XX PS Claim 1; Page 246-247; 394pp; English.
XX CC The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in AB278237-AB278293 represent cDNA encoding the
CC A. niger proteases of the invention.
XX SQ Sequence 1665 BP; 374 A; 455 C; 424 G; 412 T; 0 other;
Query Match 7.4%; Score 118.6; DB 24; Length 1665;
Best Local Similarity 47.2%; Pred. No. 1.1e-24;
Matches 673; Conservative 0; Mismatches 649; Indels 105; Gaps 6;
QY 121 GCGCTCGCGAGACTACCCGGGTGTCAAAATCCCTACTCTGTGATATGTGACACCTCTCCC 180
DB 130 GGAATTTGCAAACTACCGGGTGTCTCGGATCTTATTGGGCTATGACAGCTTCCCGCC 189
QY 181 GAGTCCCATCTTCTTCTGTCTTCTGAGCCGAGACATACCCAGAACTGACCTATC 240
DB 190 GCTTCAACAGCTTCTTGTGTTTTCGAGCCCGCAAGATCCAGCAATGCGGCTCTG 249
QY 241 ACATTGTGTGAATGTGGCCCGGAGAGCGATTCTTGTGATCGGTTCTTTCGAAGAGTTG 300
DB 250 GCCATCTGGCTCAATGGGGTCCGGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 309
QY 301 GGCCTTGCCATGTCAATTCGACTTTTCATGAC---TACATCAACCCCTCACTCGTGGAC 357

Db	1294	ACCGCCAGCTGGGCAACTACAGCTTACATCGGTCTTCCAAAGCGGGCATGAGGTCGCC	1353
QY	1432	TACTACACGACCATCGCICCCCTGCAATGTTTAAACGGACATATCTTCGTTGGGATATC	1491
Db	1354	TCTTACACGCTGTCGGGGGTATGAGATCTTCAICGGGGGACATCTCAACAAGATATC	1413
QY	1492	GCAGAGGGCAGCAAGAGATCTGCGCCAGCTCAACAGCAATGGAAAC	1538
Db	1414	CCTACTGGCCTCTTGCTGTGTGATGACGAATCCAGTCGGTTGGACC	1460
RESULT	7		
ID	ABZ78241		
XX	ABZ78241	standard; cDNA; 1581 BP.	
AC	ABZ78241:		
DT	24-APR-2003	{first entry}	
XX			
DE		A. niger serine carboxypeptidase cDNA #1.	
XX			
KW		Protease; fungal infection; aspergillosis; food; tanning; detergent;	
KW		protein solubility; viscosity; taste; texture; nutritional value;	
KW		gene; ss.	
XX			
OS		Aspergillus niger.	
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1581	
FT		/*tag= a	
FT		/tag_number= "3.4.16.6"	
XX			
PN	WO200268623-A2.		
XX			
PD	06-SEP-2002.		
XX			
PF	22-FEB-2002;	2002WO-EP01984.	
XX			
PR	23-FEB-2001;	2001EP-0200657.	
PR	23-FEB-2001;	2001EP-0200658.	
PR	23-FEB-2001;	2001EP-0200660.	
PR	26-FEB-2001;	2001EP-0200706.	
PR	26-FEB-2001;	2001EP-0200707.	
PR	26-FEB-2001;	2001EP-0200708.	
PR	26-FEB-2001;	2001EP-0200719.	
PR	28-MAR-2001;	2001EP-0000075.	
PR	28-MAR-2001;	2001EP-0000078.	
PR	28-MAR-2001;	2001EP-0000080.	
PR	28-MAR-2001;	2001EP-0000087.	
PR	28-MAR-2001;	2001EP-0000088.	
PR	21-MAY-2001;	2001EP-0000156.	
PR	21-MAY-2001;	2001EP-0000159.	
PR	21-MAY-2001;	2001EP-0000160.	
PR	21-MAY-2001;	2001EP-0000162.	
PR	21-MAY-2001;	2001EP-0000165.	
PR	21-MAY-2001;	2001EP-0000166.	
PR	21-MAY-2001;	2001EP-0000168.	
PR	21-JUN-2001;	2001EP-0000240.	
PR	21-JUN-2001;	2001EP-0000242.	
PR	21-JUN-2001;	2001EP-0000244.	
PR	21-JUN-2001;	2001EP-0000246.	
PR	12-JUL-2001;	2001EP-0000280.	
PR	12-JUL-2001;	2001EP-0000285.	
PR	30-JUL-2001;	2001EP-0000323.	
PR	30-JUL-2001;	2001EP-0000327.	
PR	02-AUG-2001;	2001EP-0000341.	
PR	02-AUG-2001;	2001EP-0000342.	
PR	02-AUG-2001;	2001EP-0000343.	
PR	02-AUG-2001;	2001EP-0000344.	
PR	09-AUG-2001;	2001EP-0000357.	
PR	16-AUG-2001;	2001EP-0000374.	
PR	16-AUG-2001;	2001EP-0000377.	
PR	20-SEP-2001;	2001EP-0000478.	

Db 1172 GTGGGAGCCGCTG-----CGTTCAGTCAAGCGCGCAACTTCAGTAGTGGG 1222
QY 1331 GTACACGCC---CNGAAGTCACAGCGGCTGAGTATGGGAACTCGGAGTATGCTA 1387
Db 1223 GTACACCAACATGTCACCTCGATGGATGACACACGCGCCAGGTGCGCGCGGGG 1282
QY 1388 AFTTCCTCTACTCCGCTGATGAGCGAGGCGCATGAAGTCCATACACGCCCATCG 1447
Db 1283 AATTTCCTCTGCGGAGTATGAGAGTGGACATGAGGTTCCTTCTATCACCTTGC 1342
QY 1448 CTCCCTGCAATGTTTAAACCGGACATCTCTCGTTGGGATATCGCAGAGGCGCAAGA 1507
Db 1343 TTCCGCTGAGAGTGTGAGCGGCTCATTCGCGCAAGGATGCGGACGGAAGATTC 1402
QY 1508 AGATCTGGCCACGCTACAGACGAAATGGACGGCTA 1543
Db 1403 CCATCTCGTGGAGTTTACAGACGCTGGCGACGCCA 1438
RESULT 8
ABZ78184
XX ID ABZ78184 standard; DNA; 2940 BP.
AC ABZ78184;
XX
DT 24-APR-2003 (first entry)
XX
DE A. niger serine carboxypeptidase gene #1.
XX
KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW gene; ds.
XX
OS Aspergillus niger.
XX
PN WO200268623-A2.
XX
PD 06-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-EP01984.
XX
PR 23-FEB-2001; 2001EP-0200657.
XX
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-MAY-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.

PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 13-NOV-2001; 2001EP-0004464.
PR 21-DEC-2001; 2001EP-0005117.
XX (STAM) DSM NV.
XX
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albang R;
XX
XX WPI: 2002-723203/78.
DR P-PSDB; ABR38817.
DR
XX Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
XX applications
XX
XX Claim 1; Page 109-111; 394pp; English.
XX
XX The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABZ78180-ABZ78236 represent genes encoding the
CC A. niger proteases of the invention.
XX
SQ Sequence 2940 BP; 691 A; 818 C; 732 G; 699 T; 0 other;
Query Match 7.3%; Score 116.8; DB 24; Length 2940;
Best local Similarity 52.5%; Pred No. 5e-24;
Matches 334; Conservative 0; Mismatches 287; Indels 15; Gaps 3;
QY 914 ACCGCACGCAATAGCTGACTACGCGCTCTGCGCGAAGCCACCAATGTGCGAGGACA 973
Db 1492 ACTGCGCGCGCGAGGATCGACGAGATCTGCGACACTGCGCGAGCTTTTTCGCGCAAG 1551
QY 974 ATGTTGAGGGCCATACACTACGCTTTGCTGGTCTGGTGTATGATATTCGCGCATCCAT 1033
Db 1552 AGGTGCAAAACGCTACGACATTTTACTCGGTGGGATGAGTATGATCTTGTGACTCA 1611
QY 1034 ATGATGACCGGACTCGCGCAAGTATTACAAACAATTTCTGGCAAGGACTGTCTATGG 1093
Db 1612 CTCCGAGCCGTTCCCTTACGAGTCTACGTTGACTACCTGACAAAGCGTCCGTCAGG 1671
QY 1094 AGGCTATCGGCGTCAACATCACTACACCGAGTCCCAATATGAGCTCTACTAGCTTCC 1153
Db 1672 CGGCATCGGCGCATACATCAATTAACAGGAGCAACACTCTTGGACTCGCTTTT 1731
QY 1154 AGCAACAGGCGACTTTGCTGGCCC---AACTCATCGAAGACTCGAGGAGATCTTGTG 1210
Db 1732 CGTCCACCGGTGACGACGGCGGACTCATGAACACCATCCAGATGTGGCAAGCTCTCA 1791
QY 1211 CTCTCCCGGTGGTGTCTCCCTCATCTATGGGACGCCGATTTACATCTGCAACTGTTCG 1270
Db 1792 AACAGGTGTCACGGTGGTTCATCTACGCGGGGATCGCGACTATAACTGCAACTGCTGG 1851

Db 1822 CTGGCGAGCTCGGAGATCCCTCATGTATGGCGACAATGAGACGGTAATAATGACTCTCG 1881
 QY 947 CCGAAGCCACCAACATGTCAGGGACAATGTTGAGGGCCATACACGCCCTTTCGIGTC 1006
 Db 1882 AGGAGCGCTCGGACTACTGTTCCGGGGAGATCAAGAGCCTGTATAGCAATACCTCCGGC 1941
 QY 1007 GTGGTGTATGATATTCGGCATCCATATGATGACCCGACTCCGCCAAGTTATACAACA 1066
 Db 1942 GAGGATACAGACATAGCGCATTTACGCCGGATGACGCTCTGTCGCTTACTTCGTCG 2001
 QY 1067 AATTTCTGCAAGAGACTCTGTATGAGCGTATCGGGGTCAACATCAACTACACCCAGT 1126
 Db 2002 GGTCTTGAATGCGCCATGTTGTCGCAAAAGGCACTTGGGGTCCCGGTGAACATACCATGT 2061
 QY 1127 CCAATATGAGCTTACTACCTTCCAGCAACAGCGGACT-----TTGCTGCG 1177
 Db 2062 CGTCAGAGCGAGTGGGGAACAGTTTCGCTCGACGGCGGATATCCGGGAATGATCCCC 2121
 QY 1178 CCACTTCAATGAGAGCTCGAGAGATTCCTCTCCCGTGGTGTCTCCCTCATCT 1237
 Db 2122 GCGAATGATGGGATATGGATACCTTGCTGACTCGGTGTCAAGGTCGATGAT 2181
 QY 1238 ATGGCGACGGATATGATGCACTGCAACTGTTTCGCGGTGACGGCGTTTCCCTCGCTCGGA 1297
 Db 2182 ATGGGACCGGACTATGCTTGTCCGTGGCGCGCGGGAGAGATGTCAGCGTGTGTTGG 2241
 QY 1298 ACTACTCCCAAGCGCCAGTTCGGAAGCGCAGGTACACGCCCTCGAAAGTCAACGGG 1357
 Db 2242 AGTACGAGGATCGGAGAAGTTCGCTGCTGCTGATGCGAGTGCAGCAGGAATCAT 2301
 QY 1358 TCGAGTATGGG---AAACTCGGAGTATGATGTAATTTCTCTTCACTCGCGTCTATGAGG 1414
 Db 2302 CCTACGTTGGGGTCTAGTAAGCAGTATGGGAACCTCTGTTACACGCTGCTTTCAGG 2361
 QY 1415 CAGGCCATGAAGTCCCATACAGCCCATCGCTCCCTCGCAITGTTAAACGGGACTA 1474
 Db 2362 CGGECATGAGTGCCTATTTATACGCCCGAAGCGGTATGAGATTTTAAATCGCGCTC 2421
 QY 1475 TCTTCGGTGGGATATCGCAGAGG 1499
 Db 2422 AGTTAATGGGATATTCGACGGG 2446

RESULT 11

AAFL1216
 ID AAFL1216 standard; cDNA; 678 BP.
 AC AAF12116;
 XX
 XX
 XX
 DE 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:4639.
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 OS Aspergillus oryzae.
 XX
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 XX 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI: 2000-594572/56.
 DR
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 88; Page 1971; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 678 BP; 153 A; 188 C; 175 G; 161 T; 1 other;
 Query Match 6.4%; Score 103; DB 21; Length 678;
 Best Local Similarity 53.0%; Pred. No. 3.2e-20;
 Matches 273; Conservative 0; Mismatches 230; Indels 12; Gaps 2;
 QY 1001 CTGCTCGTGTGTATGATATTCGCGCATCCATATGATGACCCGACTCCGCCAAGTTATT 1060
 Db 3 CTGCGAGAGCTATTACTATCTGCGATTTTACCCCGCATCTTCCGCCCTCCGTA 62
 QY 1061 ACAACAAATTCGGCAAGGACTCTGTATGACGCTATCGGTGCTCAACATCAACTACA 1120
 Db 63 ACTTTGGTTTCTTAAACACGACTGGGTCAGAGAGCCCTCGGGTTCCGGTTAAATTTCA 122
 QY 1121 CCCAGTCCATATAGTCTACTAGCTTTCCAGCAACAGCGAGCTTTTCTGTGGCCCA 1180
 Db 123 CCGAATCTGTGTATAGCGTCTACATATGGCTTTTCCGCAACAGGTGACTACCCAGCTCCG 182
 QY 1181 ACTTCATCGAAGACCTCGAGGAGATC-----CTTGCTCTCCCGTGCCTGCTCC 1231
 Db 183 ATGTACCGGTACCTGGAGGATATCGATACGTGCTTGTACTCTGGCATCAAAAGTTGCC 242
 QY 1232 TCATCTATGGCGACGCCGATTACATCTGCAACTGGTTCGCGGTTCAGCGGTTCCTCC 1291
 Db 243 TAGTGTATGCGATCGGATATGCGATGCCCTGGACGAGGAGAGAGTGAAGTGA 302
 QY 1292 CTGCGAAGTACTCCCAAGCGCCAGTTCGGAAGCGAGGTACAGCCCTCGAAAGTCA 1351
 Db 303 AGGTGAGTACTCGGATCGGCAAGTTCGCTCGGGTTACGCCCTCTGAGACCA 362
 QY 1352 ACGGCTCGAGTATGGGAACT---CGCGATATGTAATTTCTCTTCACTCGGTCT 1408
 Db 363 ATGCTCATATCTAGTGTGCTGGTCCGACAGTACGGAACATCTCGTTCACTCGGT 422
 QY 1409 ATGAGCGAGGCCATGAAGTCCCATACTACAGCCCATCCCTCCCTGCAATGTTTAA 1468
 Db 423 TCGAAGCGGTCATGAGTGGCGGATATCACCTGAAACGGGTATGATGATCTTCCACC 482


```

CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 2503 BP; 703 A; 475 C; 517 G; 808 T; 0 other;

Query Match          5.4%; Score 86.4; DB 15; Length 2503;
Best Local Similarity 57.4%; Pred. No. 6.1e-15;
Matches 156; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 145 GTCAATCCTACTCTGGATATGTGCACACCTCTCCCGAGTCCCATACCTCTCTCTGGTTC 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 624 GTGAACAATGTCGGGATATTAGATTACCAAGNCTCAAAACACTTCTTTATTGGTTT 683

QY 205 TTGGAAGCCAGACATAACCCAGAACTGCACCTATACATATGTGTGAATGGTGGCCCI 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 TTTGAGTCTAGAAATGACCCAGAGAAATGACCCAGTGATATGSHIAACGGTGTCTT 743

QY 265 GGAAGCGATTCTTGATCGGTCTCTTCGAAGAGTGTGGCCCTTGCATGTCAATTGACT 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 GCCTGTCTCTTTCGTGGGTCTTTCTTTTGAATTGGGACCTCTCTTATAGAGCTGAT 803

QY 325 TTTGATGACTACATCAACCTCACTCGTGGAGAGGTCTCCAAATTTACTATTCTGTCC 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 804 ITGAACCCATTATACACCCCTACICTTGGAAATTCACACGCTTCGTGAATTCCTAGAT 863

QY 385 CAGCCATGGAGTCGGCTTTTCATATAGTGA 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 864 CAGCCTGTGTGTGTGGTCTCTCATACGGTGA 895

RESULT 14
AAT28284
ID AAT28284 standard; cDNA; 2002 BP.
XX
AC AAT28284;
XX
DT 14-AUG-1996 (first entry)
XX
DE A. niger SFAG 2 carboxypeptidase Y cDNA.
XX
KW Carboxypeptidase Y; CPY; ascomycete; deuteromycete; host cell;
KW protease deficiency; ds.
XX
OS Aspergillus niger strain SFAG 2.
XX
FH Key Location/Qualifiers
FT CDS 140..1813
FT /*tag= a
FT mat_peptide 554..1810
FT /*tag= b
XX
PN WO9609397-A1.
XX
XX 28-MAR-1996.
XX
PF 19-SEP-1995; 95WO-US11945.
XX
PR 20-SEP-1994; 94US-0309341.
XX
PA (NOVO ) NOVO NORDISK BIOTECH INC.
XX
PI Thompson SA, Yaver DS;
XX
XX WPI; 1996-188458/19.
DR P-PSDB; AAR96738.
XX
XX Nucleic acid construct encoding a filamentous ascomycete or
PT deuteromycete carboxypeptidase Y - useful to produce host cells
PT modified to produce reduced amounts of carboxypeptidase
XX
XX Claim 5; Page 23-25; 46pp; English.
XX
CC The CPY gene (AAT28284) of Aspergillus niger strain SFAG 2 codes for
CC carboxypeptidase Y pre-propeptide (AAR96738). It was obtd. by

```

```

CC screening a cDNA library in Lambda ZAPII with a probe obtd. by PCR
CC amplification of A. niger Bo-1 genomic DNA (see also AAI28283).
CC The gene is useful in the creation of CPY-deficient mutants of
CC Aspergillus. For example, a selectable marker may be cloned into the
CC middle of the CPY gene. The resulting CPY-deficient mutant is useful
CC as a host for prodn. of heterologous protein.
XX
SQ Sequence 2002 BP; 416 A; 590 C; 506 G; 490 T; 0 other;

Query Match          5.2%; Score 84; DB 17; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.9e-14;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACICTCCCGATGCCATACCTTCTTCTGGTCTTGGAGCCAGACATAACCCAGAA 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 GACAACGAGAACGACAGCATCTGTCTACTGTGTCTTCGAGTCTCGCAATGACCCCGAG 676

QY 229 ACTGCACCTATCACATGTGTGTTGAATGTGGCCCTGGAAGGATCTTTGATCGGTCTC 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 AATGACCCCTGTGTTCTGTGGCTGAACGGTGGCCCTGGATGCTTTCCTCACCGGTCTT 736

QY 289 TTCGAAGAGTTGGGCCCTTGCCATGTCAATTGGACTTTTGTATGACIACATCAACCTCAC 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 TTCATGGAGCTCGGCCCTAGCAGCATCAACAAGAAGATCCACCCGGTCTACAACACTAC 796

QY 349 TCGTGGACAGGAGTCTCCAAITTTACTATTCCTGTCCCAAGCCATTGGGAGTCGGCTTTC 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 GCTTGGAACTCCACGGCTCGGTGATCTTCCTTGACCAAGCCGTGTCAAGTCGGTACTCT 856

QY 409 TATAGTGATACGGTTGATGGGTCCATTAAACCTGTAACTGGGGTCTGGA 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 TACAGCAACTCTGCTGTGTCAGGACACACCGTTGCTGTGCAAGGAGGTCTA 906

RESULT 15
ABZ78243
ID ABZ78243 standard; cDNA; 1611 BP.
XX
AC ABZ78243;
XX
DT 24-APR-2003 (first entry)
XX
DE A. niger carboxypeptidase Y cDNA #1.
XX
KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW gene; ss.
XX
OS Aspergillus niger.
XX
FH Key Location/Qualifiers
FT CDS 1..1611
FT /*tag= a
FT /*EC_number= "3.4.16.5"
XX
PN WO200266623-A2.
XX
XX 06-SEP-2002.
XX
XX 22-FEB-2002; 2002WO-EP01984.
XX
PR 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 15:55:43 : Search time 72.5 Seconds
(without alignments)
7040.892 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 2904
Sequence: 1 cttcagggaagtaacacggc.....gcattgccagtggtgtatg 1508

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-O=/con2.1/USPTO.spool_p/US09712338/runat_16092003_144324_14334/app_query.fasta_1.1799
-DB=A_Geneseq_15Jun03 -OFAST=fastn -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338.ecn_1.1.92.0runat_16092003_144324_14334 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2878	99.1	554	19	AAW56099	A. oryzae ATCC2038
2	2338	80.5	551	23	ABR38864	A. niger serine ca
3	980.5	33.8	623	23	ABR38865	A. niger serine ca
4	946	32.6	554	23	ABR38859	A. niger serine ca
5	914.5	31.5	526	23	ABR38817	A. niger serine ca
6	636	21.9	508	23	ABG30049	S. cerevisiae BAX-
7	630	21.7	536	23	ABR38819	A. niger carboxype
8	630	21.7	556	17	AA956737	A. niger Bo-1 carb
9	620	21.3	557	17	AA956738	A. niger SPAG 2 ca
10	568.5	19.6	491	15	AA848059	Sequence of protea
11	541.5	18.6	550	23	ABG33281	C. albicans BAX-as
12	414	14.3	481	23	ABR38845	A. niger carboxype
13	383	13.2	482	21	AAG30085	Arabidopsis thalia
14	383	13.2	502	21	AAG30064	Arabidopsis thalia
15	366.5	12.6	455	23	ABR38843	A. niger carboxype
16	361.5	12.4	479	21	AAG23905	Arabidopsis thalia
17	361.5	12.4	486	21	AAG23904	Arabidopsis thalia
18	361	12.4	476	19	AAW72966	Human serine carbo
19	361	12.4	476	22	AA88381	Human membrane or
20	359	12.4	476	20	AAV13372	Amino acid sequenc
21	359	12.4	476	21	ABQ1407	Human TANGO 176.
22	359	12.4	476	22	AAU29228	Human PRO polypept
23	359	12.4	476	22	AA88587	Human hydrophobic
24	359	12.4	476	22	AA880240	Human PRO polypept
25	359	12.4	476	24	ABU69550	Human secreted/pro
26	359	12.4	476	24	ABU71316	Human secreted/pro
27	359	12.4	476	24	ABU71473	Human PRO polypept
28	359	12.4	476	24	ABU71919	Human secreted/pro
29	359	12.4	476	24	ABU65773	Human secreted/pro
30	359	12.4	476	24	ABU66106	Human secreted/pro
31	359	12.4	476	24	ABU67373	Human secreted/pro
32	359	12.4	476	24	ABU67610	Human secreted/pro
33	359	12.4	476	24	ABU64527	Human secreted/pro
34	359	12.4	476	24	ABU65468	Human secreted/pro
35	359	12.4	476	24	ABU58604	Human PRO polypept
36	359	12.4	476	24	ABU56140	Human secreted/pro
37	359	12.4	476	24	ABU57135	Human PRO polypept
38	359	12.4	476	24	ABU54375	Human secreted/pro
39	359	12.4	476	24	ABU10714	Human secreted/pro
40	359	12.4	477	22	AAW25810	Human secreted/pro
41	357.5	12.3	479	21	AG47178	Human protein sequ
42	357.5	12.3	486	21	AG47177	Arabidopsis thalia
43	357	12.3	476	20	AA28570	Secreted peptide c
44	357	12.3	476	22	AAU39043	Human secreted pro
45	357	12.3	476	23	ABW55752	Human polypeptide

ALIGNMENTS

RESULT 1
AAW56099
ID AAW56099 standard; Protein; 554 AA.
XX
AC AAW56099;
XX
DT 27-AUG-1998 (first entry)
XX
DE A. oryzae ATCC20386 carboxypeptidase I protein.
XX
KW Carboxypeptidase I; flavour improving agent; hydrolase; proteinaceous;
KW food industry.
XX
OS Aspergillus oryzae.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= signal

FT Protein 19...554
 FT /label= carboxypeptidase I
 XX WO9814599-A1.
 XX 09-APR-1998. 102b
 XX 03-OCT-1997; 97WO-US17977.
 XX 27-NOV-1996; 96US-0757534.
 XX 04-OCT-1996; 96US-0726880.
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX (NOVO) NOVO-NORDISK AS.
 XX Berka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;
 XX Klotz A, Mathisen TE, Rey M;
 XX WPI; 1998-240098/21.
 XX N-PSDB; AAV28620.
 XX Carboxypeptidase from *Aspergillus oryzae* - hydrolyses proteinaceous
 PT substrates, useful for improving flavour of foods
 XX Claim 1; Fig 3; 82pp; English.
 CC This sequence represents carboxypeptidase I from *Aspergillus oryzae*.
 CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
 CC 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a
 CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and a
 CC 60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where
 CC N-CBZ is N-carboxy and X is any amino acid. The carboxypeptidases
 CC can be used for obtaining hydrolysates (which can be enriched in free
 CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous
 CC substrates. The carboxypeptidases can be used in flavour-improving
 CC compositions in the food industry. The products can also be used for the
 CC production of polypeptides free of carboxypeptidase activity.
 XX Sequence 554 AA;
 SQ
 Alignment Scores:
 Pred. No.: 1.43e-253 Length: 554
 Score: 2878.00 Matches: 536
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.10% Indels: 0
 DB: 19 Gaps: 0
 US-09-712-338-1_COPY_55_1662 (1-1608) x AAW56099 (1-554)
 QY 1 CTTCCAGGAAGTACACCGGCTCGCTCGTAGAAGACAGCTACCCAGAACCCACCGGG 60
 DB 19 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 38
 QY 61 GTCAAGACTCTTACACCGCAACAATGTCCACCATCCGGTACAGGAACCCGGGCGAG 120
 DB 39 VallysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu 58
 QY 121 GCGCTCGGAGACTACCCGGTGTCAATCTCTACTCTGATATGTCGACCTCTCC 180
 DB 59 GlyValCysGluThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 78
 QY 181 GAGTCCCATACCTTCTCTGCTTCTCGAACCCAGACATACCCAGAACTGCACTATC 240
 DB 79 GluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 98
 QY 241 ACATTGTGGTGAATGGTGGCCCTGGAGCGATTTCTTGTATCGGTCTCTTCGAGAGTIG 300
 DB 99 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeu 118
 QY 301 GGCCCTCCCATGCAATTCGACTTTTGTATGACTACATCAACCTCTCTCTGGAACGAG 360
 DB 119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerIrpAsnGlu 138

QY 361 GTCTCCAAITTTACTATTCTCTGCCAGCCATGGGAGTCGGCTTTTCATATAGTGATAGC 420
 DB ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 158
 QY 421 GTTGATGGCTCCATTAACCTGTAACTGGGTCGTCGAAATTCAGACTTTTCAGAGATT 480
 DB ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 178
 QY 481 CAGGGCGGTACCCAAACCATTTGATCCACTCTGATCGATACTACTACCAATCTTGGCCGAG 540
 DB GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGlu 138
 QY 541 GCGCTTGGAGATCTCTGCAAGGATTCCTTAGTGAGTACTACTAGCTTGGACTCTAGGGTG 600
 DB AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 218
 QY 601 CAGCTCAAGGACTTCAGTCTATGACGGAGAGAGTATGGAGGCGACTATGCTTCGTGATTC 660
 DB GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238
 QY 661 TTCAATCATTTTACGAGCAGAATGAGAGAAATGGCAACGGTAGTGTAAATGGTGTTCAG 720
 DB PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 258
 QY 721 CTTAATTTCAACTCTCTGGGAATTTAAGCGCATCATCGACGAGCGGATCCAGGCCCT 780
 DB LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro 278
 QY 781 TACTACCTCAATTCGCTGCGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGCGTC 840
 DB TyrTyrProGluPheAlaValAsnThrTyrGlyIleLysAlaValAsnGluThrVal 298
 QY 841 TACAACATACATGAGTTGGCCACCAATGCCAATGGTTGCCAGGATTTGATTTCCACC 900
 DB TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 318
 QY 901 TGCAACAGACAAACCCGCGATTTAGCTAGCTACGCCCTCTCGCGGAGCGACCAAC 960
 DB CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 338
 QY 961 ATGTGAGGACATGTTGAGGGCCATCTACGCCCTTTGCTGCTGCTGTGTATGAT 1020
 DB MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 358
 QY 1021 ATTCCGATCCATATGATGACCCGACTCCGCGCAAGTTATACACAATTTCTGGCAAG 1080
 DB IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 378
 QY 1081 GACTCTGTCATGGAGCTATCGGGTCAACATCAACTACCCAGTCCCAATTAATGACGTC 1140
 DB AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspVal 398
 QY 1141 TACTACGCTTCCAGCAACAGCGACTTTGCTGGCCCAACTTCATCGAAGACCTCGAG 1200
 DB TyrTyrAlaPheGlnGlnThrGlyAspPheValIrpProAsnPheIleGluAspLeuGlu 418
 QY 1201 GAGATCCTTGTCTCCCGCTCGCTCTCCCTCATCTATGCGGCGCGGATTCATCTGC 1260
 DB GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 438
 QY 1261 AACTGGTTCGGGCTCAGCGGCTTCCCTCGCTGGAACCTACTCCCAAGCGCCGACGTC 1320
 DB AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 458
 QY 1321 CGAAGCGAGGTACACGCCCTCGTAAAGTCAACGCGCTCGAGTATGCGGAACCTCGCAG 1380
 DB ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 478
 QY 1381 TATGTAATTTCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 1440
 DB TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 498


```

361 GTCCTCAATTTCTATCTCTCCAGCAGCATGGAGTCGGCTTTCATATAGTGATACG 420
139 ValThrAsnLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerGluThr 158
421 GTTGATGGTCCATTAACCTCTAATCGGGGTCGGTGAATAATTCGAGCTTTCGAGAGTT 480
159 GluAlaGlySerLeuAsnProPheThrGlyAlaValGluAsnAlaSerPheAlaGlyVal 178
481 CAGGGCCGTACCCACCAATTCATGCTACCTCTGATGCAATACCTACCAATCTTCCCGCAGAG 540
179 GlnGlyArgTyrProValIleAspAlaThrIleAspThrThrAspIleAlaAlaArg 198
541 GCGCTTGGAGATCTCTGCAAGATTCCTTAGTGACTACCTACCTAGCTTGGACTCTAGGGTG 600
199 AlaThrTrpGluValLeuGlnGlyPheLeuSerGlyLeuSerGlnLeuAspSerGluVal 218
601 CAGTCTAGGACTTCAGTCTATGAGCGGAGAGCATGATGGAGGCGACATGCTCTGCAATC 660
219 LysSerLysGluPheAsnLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238
661 TTCATCAATTTTACGAGCAGATGAGAGNAATGCCAAGCTAGTGTATGCTGTTCAG 720
239 PheAsnHisPheTyrGluGlnAsnSerLysIleAlaSerGlyGluValAsnGlyValGln 258
721 CTTAATTTCAACTCTCTGGGAATTAATTAACGGCATCATCGAGCGGATCCAGGCCCT 780
259 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspAlaIleGlnAlaAsp 278
781 TACTACCTCGAATTCGTGTGAACAATACCTACGGTATCAAGCTGTCAACGAGACCGTC 840
279 TyrTyrAlaAspPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnAspThrVal 298
841 TACAACCTACATGAGTTGTCACCAACCAATCCAAATGGTTGCCAGGATTTGATTCACC 900
299 TyrAsnTyrMetLysPheAlaAsnThrMetProAsnGlyCysGlnAspGlnValAlaSer 318
901 TGCAACAGACAAACCCAGCAGCATTCAGTCTACCTACCGCTCTCGCGCGAAGCCACCAAC 960
319 CysLysLeuThrAsnArgThrSerLeuSerAspTyrAlaIleCysThrGluAlaAlaAsn 338
961 ATGTGACGAGCAATGTTAGGGGCCATCTACGCCCTTTCGTGCTGGTGTGTATGAT 1020
339 MetCysArgAspAsnValGluGlyProTyrTyrGlnPheGlyGlyArgGlyValTyrAsp 358
1021 ATTGGCATCATCATATGATGACCCGACCTCCGCAAGTTATTAACAATAATCTTCGGCAAG 1080
359 IleArgHisProTyrAsnAspProThrProProSerTyrPheValAspTyrLeuLysLys 378
1081 GACTCTGTCATGGACGCTATGGCGTCAACATCACTACACCCAGTCCCAATTAATGACGTC 1140
379 AspSerValMetAspAlaIleGlyValAspIleAsnTyrThrGluSerSerGlyGluVal 398
1141 TACTACGCTTTCAGCAAAACAGGCGACTTGTCTGGCCCACTTCATCGAAGACTCGAG 1200
399 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 418
1201 GAGATCCTTCTCTCCCGTCCGCTCTCCCTCATCTATGCGCGCGCGATTCATCTGCG 1260
419 GluIleLeuGlnLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 438
1261 AACTGGTTCGGCGGTAGCGCGTTCCTCTCGCTGCAACTACTCTCCAGCGCGCGAGTTC 1320
439 AsnTrpPheGlyGlyGlnAlaIleSerLeuAlaValAsnTyrProHisAlaAlaGlnPhe 458
1321 CGAAGCGCAGGTCACGCGCCCTGAAGTCAACGCGTCGAGTATGGGAACATCGCGAG 1380
459 ArgAlaAlaGlyTyrThrProMetThrValAspGlyValGluTyrGlyLuthrArgGlu 478
1381 TATGTAATTTCTCCTTCACCTCGGCTCTATGAGCGAGCCCATGAGTCCCATACACGAG 1440
479 TyrGlyAsnPheSerPheThrArgValTyrGlnAlaGlyHisGluValProTyrTyrGln 498
1441 CCCATCGCCTCCCTGCAATGTTTAACCGGACATCTTCGGTGGGATATCCGAGAGGC 1500
```

```

499 ProIleAlaAlaLeuGlnLeuPheAsnArgThrLeuPheGlyTrpAspIleAlaAlaGly 518
1501 CAGAAAGAATCTGGCCAGCTACAGAGCAATGAAGGCTACAGCTAGCATACACAG 1560
519 ThrThrGlnIleTrpProGluTyrSerThrAsnGlyThrSerGlnAlaThrHisThrGlu 538
1561 TCGTCCGTGCGCTGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT 1590
539 SerPheValProLeuSerThrAlaSerSer 548

RESULT 3
ABR38865
ID ABR38865 standard; Protein; 523 AA.
XX AC ABR38865;
XX DE
XX DT 24-APR-2003 (first entry)
XX DE
XX KW A. niger serine carboxypeptidase polypeptide #4.
XX KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
XX KW protein solubility; viscosity; taste; texture; nutritional value;
XX KW EC3.4.16.6.
XX OS Aspergillus niger.
XX PN WO200268623-A2.
XX PD 06-SEP-2002.
XX PF 22-FEB-2002; 2002WO-EP01984.
XX PR 23-FEB-2001; 2001EP-0200657.
XX PR 23-FEB-2001; 2001EP-0200658.
XX PR 23-FEB-2001; 2001EP-0200660.
XX PR 26-FEB-2001; 2001EP-0200706.
XX PR 26-FEB-2001; 2001EP-0200707.
XX PR 26-FEB-2001; 2001EP-0200708.
XX PR 26-FEB-2001; 2001EP-0200719.
XX PR 26-MAR-2001; 2001EP-0000075.
XX PR 28-MAR-2001; 2001EP-0000078.
XX PR 28-MAR-2001; 2001EP-0000080.
XX PR 28-MAR-2001; 2001EP-0000087.
XX PR 28-MAR-2001; 2001EP-0000088.
XX PR 21-MAY-2001; 2001EP-0000156.
XX PR 21-MAY-2001; 2001EP-0000159.
XX PR 21-MAY-2001; 2001EP-0000160.
XX PR 21-MAY-2001; 2001EP-0000162.
XX PR 21-MAY-2001; 2001EP-0000165.
XX PR 21-MAY-2001; 2001EP-0000166.
XX PR 21-MAY-2001; 2001EP-0000168.
XX PR 21-JUN-2001; 2001EP-0000240.
XX PR 21-JUN-2001; 2001EP-0000242.
XX PR 21-JUN-2001; 2001EP-0000244.
XX PR 21-JUN-2001; 2001EP-0000246.
XX PR 12-JUL-2001; 2001EP-0000280.
XX PR 12-JUL-2001; 2001EP-0000285.
XX PR 30-JUL-2001; 2001EP-0000323.
XX PR 30-JUL-2001; 2001EP-0000327.
XX PR 02-AUG-2001; 2001EP-0000341.
XX PR 02-AUG-2001; 2001EP-0000342.
XX PR 02-AUG-2001; 2001EP-0000343.
XX PR 02-AUG-2001; 2001EP-0000344.
XX PR 09-AUG-2001; 2001EP-0000357.
XX PR 16-AUG-2001; 2001EP-0000374.
XX PR 16-AUG-2001; 2001EP-0000377.
XX PR 20-SEP-2001; 2001EP-0000478.
XX PR 20-SEP-2001; 2001EP-0000483.
XX PR 22-OCT-2001; 2001EP-0000552.
XX PR 22-OCT-2001; 2001EP-0000553.
XX PR 22-OCT-2001; 2001EP-0000554.
XX PR 22-OCT-2001; 2001EP-0000556.
```


PR 22-OCT-2001; 2001EP-0000557.
 PR 22-OCT-2001; 2001EP-0000558.
 PR 15-NOV-2001; 2001EP-0004564.
 PR 21-DEC-2001; 2001EP-0005117.
 XX (STAM) DSM NV.
 PA
 XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
 PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
 PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
 PI Stiebler J, Albarg R;
 XX
 DR WPI; 2002-723203/78.
 DR P-PSDB; ABZ78232, ABZ78289.
 XX
 PT Novel isolated protease polypeptide useful in laboratory, clinical,
 PT pharmaceutical, chemical, diagnostic, personal care and industrial
 PT applications -
 XX
 PS Claim 13; Page 382-385; 394pp; English.
 XX
 CC The invention relates to a novel isolated protease polypeptide. A
 CC polypeptide or polynucleotide of the invention is useful for diagnosing a
 CC fungal infection such as aspergillosis, or as a query sequence to perform
 CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing, in
 CC cheese manufacture, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide made in the
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABR3813-ABR38869 represent the A. niger proteases of
 CC the invention.
 XX
 SQ Sequence 623 AA;
 Alignment Scores:
 Pred. No.: 2,14e-80 Length: 623
 Score: 980.50 Matches: 216
 Percent Similarity: 56.23% Conservative: 91
 Best Local Similarity: 39.56% Mismatches: 178
 Query Match: 61
 DB: 23 Gaps: 19
 US-09-712-338-1_COPY_55_1662 (1-1608) x ABR38865 (1-623)
 QY 76 ACCGCAACAATGTCACCATCCGGTACAGGAA-----CCCGGGGCAGAG----- 120
 DB 32 ThrProGluAspLeuThrValIleHisSerGluIlePheProGlyAlaArgIleSerTyr 51
 QY 121 -----GGCGTCTGCGAGACTACCGGGGTGCAATCCTACTCTCGATATGTC 168
 DB 52 LysGlnProLeuGlyIleCysThrThrProSerThrProSerTyrSerGlyTyrIle 71
 QY 169 GACACCTCTCCGAGTCC----- 186
 DB 72 HisLeuProHisThrLeuThrAsnLeuSerIleProGlyIleSerIleSerGlnPro 91
 QY 187 -----CATACCTTCTCTGGTCTCGAAGCCAGA-----CATACCCAGAAACT 231
 DB 92 TyrProIleAsnThrPhePheTrpThrPheProSerArgHisHisAsnAsnAspThr 111
 QY 232 GCACCTATCATGTGTGAATGTCGCGCTCGAAGCGATTCTTTGATCGTCTCTC 291
 DB 112 SerProLeuThrIleTrpMetAsnGlyGlyProGlyGlySerSerMetIleGlyLeuPhe 131
 QY 292 GAAGGTGGGCCCTTGCATCTCAATTCGACTTTTGTGATGAC---TACATCAACCCICAC 348
 DB 132 GlnGluAsnGlyProCysThrValAsnThrAspSerAsnSerThrAlaTyrAsnProTrp 151
 QY 349 TCGTGGACGAGGCTCCAAATTACTATTCTGCTCCAGCCATTGGGAGTGGCTTTTCA 408

DB 152 SerTrpAsnGluTyrValAspMetLeuTyrIleGluGlnProValGlnThrGlyPheSer 171
 QY 409 TATAGTGATACCGTGTGATGGTCCATTAAACCCCTGTAACCTGGGTGGTGGTTCGAGC 468
 DB 172 TyrAspValLeuArgAsnGlyThrLeuAsp-----LeuAsnGluThr 185
 QY 469 TTTCAGGAGTTCAGGCGCGGTACCCACCACTTGTAT-----GCACCTCTGATCATGACT 522
 DB 186 Phe-----LeuValGlyThrLeuProSerGlnAspValHisGlyThrValAsnGlyThr 203
 QY 523 ACCAATCTTCCGCGAGAGCGCTGGGAGATCTCTGCAAGGATTCCTTTAGTGGGA----- 576
 DB 204 ValAsn---GlyGlyArgAlaLeuTrpValAlaLeuGlnValTrpLeuGlyGluPheSer 222
 QY 577 -----CTACCTAGCTTGGACTCTAGGTGCAGTCTAAGGAC---TTCAGTCTATGACG 627
 DB 223 GluTyrValSerSerValAspGlyAsnGlyGlyAspAspArgValSerIleTrpThr 242
 QY 628 GAGAGCTATGAGGCGCACTATGTCCTGCAFTCTTCAATCATTTTACGAGCAGAAATGAG 687
 DB 243 GluSerTyrGlyGlyArgTyrGlyProAlaIleThrAlaLeuPheGlnGluMetAsnGlu 262
 QY 688 AGAATTCACCAACGCTAGTGTAAAT---GGTGTTCAGCTTAAATTCACCTCTCTGGGAAT 744
 DB 263 ArgIleGluSerGlyGluValSerThrGlyLysIleHisLeuAspThrLeuGlyIle 282
 QY 745 ATTAACGGCATCATCGAGCGGATCCAGCGCCCTTACTACCTGATGATTCGCTGTGAAC 804
 DB 283 IleAsnGlyCysValAspLeuLeuValGlnValProSerPheProGluAlaTyrAsn 302
 QY 805 AATACCTACGGTATCAAGGCTGTCAACGAGACCGCTCTACCACTACATGAAGTTGCCAAC 864
 DB 303 AsnThrTyrGlyIleGluGlyIleAsnArgThrLeuTyrAspArgAlaMetAspSerTrp 322
 QY 865 CAATGCCAATGTTGCCAGGATTTGATTTCCACCTGCAACAGCAACACCGCACCGCA 924
 DB 323 SerLysProGlyGlyCysArgAspMetIleGluCysAspAspAlaGlyGlu----- 340
 QY 925 TTACGTGAC---TACGCCCTCTCGCGGAGCCAGCCACCAACATGTGCGAGGCAATGTGAG 981
 DB 341 LeuGlyAspProLeuIleCysGluGluAlaSerAspTrpCysSerArgGluIleLys 360
 QY 982 GGCCCATATACCGCTTGTGCTGTGTGTGTATGATATTCGATTCGCAATATGATGAC 1041
 DB 361 SerLeuTyrThrAsnThrSerGlyArgGlyTyrIleAspIleAlaHisPheThrProAsp 380
 QY 1042 CGGACTCCGCGCAAGTATTACACAAATTTCTGCGCAAGGACTCTGTCATGGACGCTATC 1101
 DB 381 AlaAlaLeuValProTyrPheValGlyPheLeuAsnArgProTrpValGlnLysAlaLeu 400
 QY 1102 GCGCTCAACATCACTACACCCAGTCCAATAATAGCTCTACTACGCTTTCAGCAACAA 1161
 DB 401 GlyValProValAsnTyrThrMetSerSerGluAlaValGlyAsnSerPheAlaSerThr 420
 QY 1162 GCGGCTTGTCTGCGCC-----AACTTCAICGACCTCTCGAGGAGATC 1206
 DB 421 GlyAspTyr-----ProArgAsnAspProArgGlyMetIleGlyAspIleGlyTyrLeu 438
 QY 1207 CTGTCTCTCCCGTGGTGTCTCCCTCATCTATGGGAGCGCGGATTCATCTGCAACTGG 1266
 DB 439 LeuAspSerGlyLysValAlaMetValTyrGlyIleAspArgAspTyrAlaCysProTrp 458
 QY 1267 TTCGCGGTGAGCGGCTTTCCTCGTCCGAACTACTCCAAAGCCCGCCAGTCCGAGCC 1326
 DB 459 ArgGlyGlyGluAspValSerLeuLeuValGluTyrGluAspAlaGluLysPheArgAla 478
 QY 1327 GCAGGTTACGCGCCCTGAAAGTCACCGCGCTCGAGTAT-----GGGAAACTCGCGAG 1380
 DB 479 AlaGlyTyrAlaGluValGlnThrLysSer---SerTyrValGlyGlyLeuValArgGln 497
 QY 1381 TATGGTAATTCTCTCTCCTCAGTGGCTCTATGAGGAGCGCAAGAAAGTCCCATCTACAG 1440

498 TyrGlyAsnPheSerPheThrArgValPheGlnAlaGlyHisGluValProPheTyrGln 517
1441 CCCATCGCTCCCTGCATTTCTTAACCGGACTATCTTCGTTGGGATATCGCAGAGGCG 1500
518 ProGluThrAlaTyrGluLeuPheAsnArgAlaGlnPheAsnTrpAspIleAlaThrGly 537
1501 CAGAAAGAGACTGG-----CCACAGCTACAGAGCAATGGAAACGGCTACAGCTACGCAT 1554
538 GlyIleSerLeuGluGlnAsnGlnSerTyrGlyThrGluGlyProSerSerThrTrpHis 557
1555 ACACAGTCGTCGTCGCCG 1572
558 IleLysAsnGluValPro 563
RESULT 4
ID ABR38859 standard; Protein: 554 AA.
AC ABR38859;
DT 24-APR-2003 (first entry)
DE A. niger serine carboxypeptidase polypeptide #2.
XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW EC3.4.16.6.
XX Aspergillus niger.
OS WO200268623-A2.
PN 06-SEP-2002.
PD 22-FEB-2002; 2002WO-EF01984.
PF 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200650.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.

22-OCT-2001; 2001EP-0000553.
22-OCT-2001; 2001EP-0000554.
22-OCT-2001; 2001EP-0000556.
22-OCT-2001; 2001EP-0000557.
22-OCT-2001; 2001EP-0000558.
15-NOV-2001; 2001EP-0004464.
21-DEC-2001; 2001EP-0005117.
(STAM) DSM NV.
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albang R;
XX WPI: 2002-723203/78.
DR P-PSDB; ABZ78226, ABZ78283.
DR Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications -
XX Claim 13; Page 366-369; 394pp; English.
PS The invention relates to a novel isolated protease polypeptide. A
XX polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABR38813-ABR38869 represent the A. niger proteases of
CC the invention.
XX Sequence 554 AA;
SQ Alignment Scores:
Pred. No.: 2.84e-77 Length: 554
Score: 946.00 Matches: 213
Percent Similarity: 51.72% Conservative: 73
Best Local Similarity: 38.52% Mismatches: 176
Query Match: 32.58% Indels: 91
DB: Gaps: 13
US-09-712-338-1_COPY_55_1662 (1-1608) x ABR38859 (1-554)
QY 37 CAGCTACCCAGAAACCCACCGGGTCAAGACTCTTACACCGCA-----ACAAATGTC 90
DQ 17 GlnPheProGluProGluGlyThrValLeuLysSerLysLeuHisGluAsnVal 36
QY 91 ACCATCGGTACAGGACCGCGGCGAGCGGCTCTCGAGACTACCCCGGGTGTCAA 150
DQ 37 ThrileSerPheLysGluPro-----GlyIleCysGluThrThrProGlyValarg 53
QY 151 TCCTACTCTGATATGTGCACACCTCTCCCGAGTCCCACTCTCTCTGTTCTCGAA 210
DQ 54 SerTyrSerGlyTyrValHisLeuProAlaSerThrSerPheThrPhePheGlu 73
QY 211 GCCAGACATAACCCAGAAACCTGCACCTATACATGTGTTGATGGTGGCCCTGGGAGC 270
DQ 74 AlaArgLysAspProSerAsnAlaProLeuAlaIleTrpLeuAsnGlyGlyProGlyGly 93
QY 271 GATCTTTGATCGGTCTCTTCGAGAGTTGGCCCTTCCCATGTCAATTCG---ACTTTT 327
DQ 94 SerSerLeuMetGlyLeuLeuGluGlyProCysSerIleAlaSerAspSerLys 113
QY 328 GATGACTACATCAACCCCTCACTCGTGTGGAACGAGGTCTCCAAATTTACTATTCTGTCGAC 387

Db 114 ThrThrValLeuAsnProTrpSerTrpAsnAsnGluValAsnLeuLeuPheLeuAspGln 133
 QY 388 CCATGGAGTGGCTTTTTCATATAGTATGATACGGTGTGATGGGTCATTAACCCGTGACT 447
 Db 134 ProThrGlnValGlyPheSerTyrAspValProThrAsnGlyThrLeuThr----- 150
 QY 448 GGGGTCTCGAAATTCGAGCTTTGCGAGGAGTTTCAGGCGCGGTACCAACCAATGATGCC 507
 Db 150 ----- 150
 QY 508 ACTCTGATGATACCTACCAATCTTGGCGAGAGCCCTTGGGAGATCCCTGCAAGGATTC 567
 Db 151 -----AlaAsnGlyThrAlaPheAlaAlaHisAlaLeuTrpHisPheAlaGlnThrTrp 168
 QY 568 CTTAGTGGACTACCTAGCTTG-----GACTCTAGGCTGAGTCTAAGGACTTTCAGT 618
 Db 169 PhePheGluPheProHisTyrLysProAsnAspArgVal-----Ser 183
 QY 619 CTATGGACGAGAGCTATGGAGGCGACTATGGTCTGTCATCTCTCAATCAATTTTACGAG 678
 Db 184 LeuTrpAlaGluSerTyrGlyGlyHisTyrGlyProGlyTlePheArgPhePheGlnGln 203
 QY 679 CAGATGAGAGAATTCGCAACGCTAGTGTI---AATGGTCTICAG---CTTAATTTCAAC 732
 Db 204 GlnAsnAspLysIleAlaGluGlyThrAlaGluAspGlyAlaGlnTyrLeuHisLeuAsp 223
 QY 733 TCCTCTGGGAATTAATACCGCATCATCGAGCGGATCCAGCGCCCTTACTACCTGAA 792
 Db 224 ThrLeuGlyIleValAsnGlyLeuMetAspMetValIleGlnGluAlaTyrIleThr 243
 QY 793 TTCGGCTGTACAAATACCTACGGTATCAAGGCTGTCAAGGAGACCGTCTACACTACAT 852
 Db 244 TrpPro-----TyrAsnAsnVal 249
 QY 853 AAGTTTGCCACCAATGCAATGCTGCCAGGATTTGATTCCACCTGCCAACAGACA 912
 Db 250 ArgLeuAlaPro-----SerSerPhe 256
 QY 913 AACCGACCGCAATAGCTGACTACGCGCTCTGCGCGAAGCCACCAACATGTCCAGGAC 972
 Db 257 AsnSerArgGlyPheArgAspGlnAlaLeuAlaCysGluAlaLeuLysGluArgAsp 276
 QY 973 -----AATGTT----- 978
 Db 277 SerGlyLeuProHisSerGlyLysAsnIleSerGluIleCysGlyGlyLeuAlaLeuGlu 296
 QY 979 -----GAGGGCCA-----TACTAGCGCTTGTGCTGCTGCTGTGTATGATATCGG 1026
 Db 297 TrpGlyAspGlyProIleThrTyrThrHisThrPheAsnArgGlyTrpTyrAspIleAla 316
 QY 1027 CATCATATGATGACCGGACTCCGCCAAGTATTATACAAATTTCTGGCAAGGACTCT 1086
 Db 317 HisProLysAsnAspProPheProAlaLysHisMetLeuGlyTyrLeuThrGlnGluSer 336
 QY 1087 GTCATGGAGCTATCGCGCTCAACACTCACTACACCGAGTCCCAATATGAGCTCTACTAC 1146
 Db 337 ValLeuAlaAlaLeuGlyValProValAsnPheThrSerSerSerAlaValAlaThr 356
 QY 1147 GCTTTCAGCAACAGCGGCTTCTCTGCGCCCACTTCATCGAAGACCTCCAGAGATC 1206
 Db 357 GlnPheIleLysThrPheAspIleValHisGlyGlyPheLeuAspAlaIleGlyTyrLeu 376
 QY 1207 CTGCTCCCGCGTGTCTCCCTCATCTATGCGGACGCGGATTCATCTGCAACTGG 1266
 Db 377 LeuAspSerGlyValLysValHisMetMetTyrGlyAspArgAspTyrAlaCysAsnTrp 396
 QY 1267 TTCGGCGTACGGGCTTTCCTCGCTCGCAACTACTCCCAAGCGCGCGAGTCCGGAAGC 1326
 Db 397 ValGlyGlyGluLysAlaSerLeuAlaValProTyrSerArgIleThrGluPheAlaAsp 416
 QY 1327 GCAGGTACAGCGCCCTCAAGTCAACGGCTCGAGTATGGGAAACCTCGGAGTATGTT 1386
 Db 417 ThrGlyTyrProLeuLeuThrProAspGlyIleSerGlyMetThrArgGlnLeuGly 436

QY 1387 AATTTCTCTCTACTCGGCTCTATGAGCAGCGCCATAGTCCCTACTACTACGCGCCATC 1446
 Db 437 AsnTyrSerPheThrArgValPheGlnAlaGlyHisGluValProSerTyrGlnProVal 456
 QY 1447 GCCTCCCTGCAATTTGTTTAAACCGGACTATCTTCGTTGGGATATGCGAGGCGCAGAAG 1506
 Db 457 AlaAlaTyrGluIlePheMetArgAlaThrPheAsnLysAspIleProThrGlyLeuLeu 476
 QY 1507 AAGATCTGGCCAGCTACAGCAATGGAACGGCTACAGCTACGCTACACATACAGTCGTCC 1566
 Db 477 AlaValAspGluPheGlnSerValGlyProLysAspThrTrpHisIleLysAsnIle 496
 QY 1567 GTGCCGCTGCTACCGGTACCGCATGTCACAGTGTG 1603
 Db 497 ProProIle-MetProLysProGlnCysTyrValLeu 508
 RESULT 5
 ABR38817
 ID ABR38817 standard; Protein; 526 AA.
 XX ABR38817;
 AC ABR38817;
 XX 24-APR-2003 (first entry)
 DT
 XX
 DE A. niger serine carboxypeptidase polypeptide #1.
 XX
 KW Protease; fungal infection; aspergillosis; food; tanning; detergent;
 KW protein solubility; viscosity; taste; texture; nutritional value;
 KW EC3.4.16.6.
 XX
 OS Aspergillus niger.
 XX
 PN W0200268623-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 22-FEB-2002; 2002WO-EP01984.
 XX
 PR 23-FEB-2001; 2001EP-0200657.
 PR 23-FEB-2001; 2001EP-0200658.
 PR 26-FEB-2001; 2001EP-0200660.
 PR 26-FEB-2001; 2001EP-0200706.
 PR 26-FEB-2001; 2001EP-0200707.
 PR 26-FEB-2001; 2001EP-0200708.
 PR 26-FEB-2001; 2001EP-0200719.
 PR 28-MAR-2001; 2001EP-0000075.
 PR 28-MAR-2001; 2001EP-0000078.
 PR 28-MAR-2001; 2001EP-0000080.
 PR 28-MAR-2001; 2001EP-0000087.
 PR 28-MAR-2001; 2001EP-0000088.
 PR 21-MAY-2001; 2001EP-0000156.
 PR 21-MAY-2001; 2001EP-0000159.
 PR 21-MAY-2001; 2001EP-0000160.
 PR 21-MAY-2001; 2001EP-0000162.
 PR 21-MAY-2001; 2001EP-0000165.
 PR 21-MAY-2001; 2001EP-0000166.
 PR 21-MAY-2001; 2001EP-0000168.
 PR 21-JUN-2001; 2001EP-0000240.
 PR 21-JUN-2001; 2001EP-0000242.
 PR 21-JUN-2001; 2001EP-0000244.
 PR 21-JUN-2001; 2001EP-0000246.
 PR 12-JUL-2001; 2001EP-0000280.
 PR 12-JUL-2001; 2001EP-0000285.
 PR 30-JUL-2001; 2001EP-0000323.
 PR 30-JUL-2001; 2001EP-0000327.
 PR 02-AUG-2001; 2001EP-0000341.
 PR 02-AUG-2001; 2001EP-0000342.
 PR 02-AUG-2001; 2001EP-0000343.
 PR 02-AUG-2001; 2001EP-0000344.
 PR 09-AUG-2001; 2001EP-0000357.
 PR 16-AUG-2001; 2001EP-0000374.
 PR 16-AUG-2001; 2001EP-0000377.

PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 15-NOV-2001; 2001EP-0004464.
PR 21-DEC-2001; 2001EP-0005117.
XX (STAM) DSM NV.
XX
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Kludbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albang R;
XX
XX WPI: 2002-723203/78.
DR P-PSDB; AB278184, AB278241.
DR
XX
XX Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications
XX
XX Claim 13; Page 271-274; 394pp; English.
XX
XX The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABR38313-ABR38869 represent the A. niger proteases of
CC the invention.
XX
XX Sequence 526 AA;
SQ

Alignment Scores:
Pred. No.: 2.08e-74 Length: 526
Score: 914.50 Matches: 211
Percent Similarity: 51.98% Conservative: 78
Best Local Similarity: 37.95% Mismatches: 186
Query Match: 31.49% Indels: 81
DB: 23 Gaps: 13

US-09-712-338-1_COPY_55_1662 (1-1608) x ABR38817 (1-526)
QY 25 GTCGGTAGAGACAGCTACCCAGAACCCACCGGGGTCAAGACTCTTACAAACCGCAAC 84
DB 16 ValSerArgAlaGlnPheValAlaProProThrAspLeuIleProThrLysGlyTyrLeu 35
QY 85 AATGTCACCATCCGGTACAGAACCCGGGCGACAGGGCGTCTGGAGACTACCCCGGT 144
DB 36 AspIleProValArgTyrLysGlnVal---ProThrGlyIleCysGluThrAspProSer 54
QY 145 GTCAATCCTACTCTGGATATGTCGACACCTCTCCCGAGTCCCATACCTCTCTGGTTC 204
DB 55 ValLysSerPheSerGlyTyrValAspValAlaGluHisGluHisIlePhePheTrpPhe 74
QY 205 TTCGAAGCCAGACAT---AAACCCAGAACTGCACCTATCAGATGTGGTTGAATGGT--- 258
DB 75 PheGluAlaArgAsnGlnAspProThrGluAlaProLeuThrValTrpIleAsnGlyGly 94
QY 259 -----GGCCCTGGAGGACATCTCTTGATCGTCTCTCGAAGAGTGGCGCCT 306
DB 95 MetSerAspProGlyProGlySerSerMetIleGlyLeuPheGlnGluHisGlyPro 114

QY 307 TGCCATGTCAATTCGACTTTTGTGACTACATCAACCCCTCAGTCGTGGACAGGCTGCC 366
DB 115 CysGlyIleAspAlaAsnGlySerValTyrAsnAsnProTyrSerTyrPheAsnAlaSer 134
QY 367 AATTTACTATTCTCTGCCAGCCATGCGGCTTTTATATAGTATAGTACGATGAT 426
DB 135 AsnMetLeuTyrIleAspGlnProValGlnThrGlyPheSerTyrSer----- 150
QY 427 GGGTCCATTAAACCTGTAACCTGGGTCGTCGAAATTCGAGCTTTCCAGGAGTTCAGGC 486
DB 151 -----IleProValProGlyTyrValAspSerSerThrAspAsnGlyPheMetGly 167
QY 487 CGSTACCCAAACCATTCATGCCACTCTGATGATACCTACCAATCTTCCCGCAGAGCGCT 546
DB 168 AlaPhePro----- 170
QY 547 TGGGAGATCCTGCAAGGATTCCTTAGTGACTACCTAGCTTAGCTTAGGTCAGTCT 606
DB 171 -----GlnTyrSerArg 174
QY 607 AAGGACTTTCAGTCTATGGACGAGAGTATGGAGGACATGATGCTCTGCAATTCAT 656
DB 175 GluThrPheHisPheThrThrGluSerTyrGlyHisTyrGlyProValPheAsnGlu 194
QY 667 CATTTTACGAGCAGATGAGAAATGCCAACGGTAGCTGTTAATGCTGTTACGCTAAT 726
DB 195 TyrIleGluGlnGlnAsnAlaHisLeuGlnProGlyAla-----LysLysIleGln 211
QY 727 TTCACACTCTCGGAATATTAACGGCATCATCAGCAGGCGATCCAGGCGCCCTACTAC 785
DB 212 LeuGlySerValMetIleGlyAsnGlyTrpTyrAspProIleGlnTyrGlnAlaTyr 231
QY 787 CCTGAATTCGCTGTG-----AACATACCTACCGTATCAAGGCTGTCAAGACAGCGTC 840
DB 232 TyrAsnPheThrValTyrProGlyAsnThrTyrAspTyrLeuProPheAsnLysSerIle 251
QY 841 TACAACATACATGAAGTTTCCACCAATGTCGAGGATGTCGAGGATTTTCATTTCCACC 900
DB 252 SerSerLeuMetTyrAsnAsnLeuTyrGlyProGlyAsnCysLeuAspGlnLeuTyrAsp 271
QY 901 TGCAACACAGACAAACCGCATTAGCTAGCTACGCTCTCTCGGCGAGGCGCAACCAAC 960
DB 272 Cys-----AlaAlaArgGlyIleAspGluIleCysSerThrAlaAspAsp 286
QY 961 ATGTCCAGGACAAATGTGAGGGCCATACATACGCTTGTGCTGTGCTGTGTATGAT 1020
DB 287 PheCysAlaAsnGluValGluAsnValTyrAspIleTyrSerGlyArgAspGluTyrAsp 306
QY 1021 ATTCGGCATCCATATGATGACCCGACTCCGCAAGTATTAACAATAATTTCTGGCAAG 1080
DB 307 PheArgGluLeuThrProAspProPheProTyrGluPheTyrValAspTyrLeuAsnLys 326
QY 1081 GACTCTGTCATGAGCTATCGGGTCAACATCAACTACACCCAGTCCCAATATGACGTC 1140
DB 327 AlaSerValGlnAlaAlaIleGlyAlaTyrIleAsnTyrThrGluSerAsnAlaVal 346
QY 1141 TACTACGCTTCCAGCAACAGGC---GACTTTGTCTGCCCACTTATCGAAGACCTC 1197
DB 347 GlyLeuAlaPheSerSerThrGlyAspGlyArgLeuMetAsnThrIleGlnAspVal 366
QY 1198 GAGGAGATCCTTGTCTCTCCCGCTGCTCTCCCTCATCTATGGGCGAGCGGATACATC 1257
DB 367 GlyLysLeuLeuLysGlnGlyValThrValValMetTyrAlaGlyAspAlaAspTyrAsn 386
QY 1258 TGCAACTGGTTCGGGGTCCAGCGCTTCCCTCGCTGCGGAACTACTCCCAAGCCGCCAG 1317
DB 387 CysAsnTrpLeuGlyGlyGluAlaValSerLeuGlnVal-----LysAlaAlaAsn 403
QY 1318 TTCGGAAGCGGAGGTACAGCCCTCGAAAGTC---AACGGCTCGATGATGGGAAGCT 1374
DB 404 PheSerSerAlaGlyTyrThrAsnIleValThrSerAspGlyValThrHisGlyGlnVal 423

QY 766 GGGATCCAGGCGCCCTTACTACCTGAATTCGCTGTG---AACAAATACCTACCGTATCAAG 822
|||||
Db 256 LeuIleGlnAlaAspTyrGluProMetAlaCysGlyLeuGlyTyrHisProVal 275
823 GCTGTCAAGGAGAGCGCTTACAACTACATGAAGATTGGCAACCAAAATGCTTGC 882
276 LeuSerSerGluGluCysGluLysMetSerLysAlaAlaGlyArg----- 290
883 CAGGATTGATTTCCACCTGCAACAGACAAACCGCACCGCATCTACTACACGCCCTC 942
291 -----CysArgArgLeuAsnLysLeuCysTyrAlaSerLysSerSer 304
943 -----TGCGCCGAGCCACCACTGTCAGGACAAATGTTGAGGGGCATACTACGCC 936
305 LeuProCysLleValAlaThrAlaTyrCysAspSerAlaLeuLeuGluProTyrIleAsn 324
997 TTGCTGCTGGTGGTGTGATGATTCGGCAATTCGCAATGATGATGACCGG----- 1044
325 ---ThrGlyLeuAsnValTyrAspIleArgGlyProCysGluAspAsnSerThrAspGly 343
1045 -----ACTCCGCCAAGTTATTACAAACAATTTCTGCARAGGACTCTGTCATGGAC 1095
344 MetCysTyrThrGlyLeuArgTyrValAspGlnTyrMetAsnPheProGluValGlnGlu 363
1096 GCTATCGGCTCAACATC---AACTACACCCAGTCCCAATAATGACGCTCTACTACGGTTTC 1152
364 ThrLeuGlySerAspValHisAsnTyrSerGlyCysAspAsnValPheThrGlyPhe 383
1153 CAGCAACAGGCGACTTGTCTGGCCCACTTCATCGACACCTCGAGGAGATCCTTGCT 1212
384 LeuPheThrGlyAspGlySerLysPro---PheGlnGlnTyrIleAlaGluLeuLeuAsn 402
1213 CTCCCGTGGTCTCTCCCTCATCTATGGCGAGCGCATACATCTGCAACTGTTGCGG 1272
403 HisAsnIleProValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGly 422
1273 GGTACGCGCTTCCCTCGCTGCGAATCTCCCAAGCCCGCCAGCTTCGGAAGCGAGG 1332
423 AsnHisAlaTrpSerAsnGluLeuGluTrpIleAsnLysArgArgTyrGlnArgMet 442
1333 TACACGCGCC-----CTGAAGATCAACGGCGTGCAGTATGGGAAATCGCGAGTATGGT 1386
443 LeuArgProTrpValSerLysGluThrGlyGlnGluLeuGlyGlnValLysAsnTyrGly 462
1387 AATTCCTCTCTACCTCCGCTATGAGCGAGCGCATGAGGCCCATCTACCAAGCGCATC 1446
463 ProPheThrPheLeuArgIleTyrAspAlaGlyHisMetValProTyrAspGlnProGlu 482
1447 GCCTCCCTGCAATGTTTAACCGGACTATCTCGT 1482
483 AlaSerLeuGluMetValAsnSerTrpIleSerGly 494

RESULT 7

ABR38819
ID ABR38819 standard; Protein: 536 AA.

AC ABR38819;

DT 24-APR-2003 (first entry)

XX A. niger carboxypeptidase Y polypeptide #1.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW EC3.4.16.5.

XX Aspergillus niger.

XX WO200268623-A2.

XX 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

XX 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 22-OCT-2001; 2001EP-0000558.
PR 15-NOV-2001; 2001EP-0000464.
PR 21-DEC-2001; 2001EP-00005117.
(STAM) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
Stiebeler J, Albang R;

WPI: 2002-723203/78.
P-PSDB; ABZ78186, ABZ78243.

Novel isolated protease polypeptide useful in laboratory, clinical,
pharmaceutical, chemical, diagnostic, personal care and industrial
applications -

Claim 13; Page 279-281; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A
polypeptide or polynucleotide of the invention is useful for diagnosing a
fungal infection such as aspergillosis, or as a query sequence to perform
a search against public databases. A polypeptide of the invention is
useful in a selected number of industrial or pharmaceutical processes, in
laboratory or clinical processes, in food industry (baking, brewing,
cheese manufacturing, meat tenderising), in tanning industry and in the
manufacture of biological detergents. A polypeptide may also be useful
for improving protein solubility, extraction yields, viscosity or taste,
texture, nutritional value, minimising of antigenicity or
anti-nutritional factors, colour or functionality as well as processing

[illegible]

PI Nucleic acid construct encoding a filamentous ascomycete or
 PT deuteromycete carboxypeptidase Y - useful to produce host cells
 PT modified to produce reduced amounts of carboxypeptidase
 XX
 XX
 XX Claim 4; Fig 1A-E; 46pp; English.
 XX Carboxypeptidase Y (AAR96737), a vacuolar protease, is the product of
 CC the CPY gene (AAR28283) of *Aspergillus niger* strain Bo-1. Creation
 CC of CPY-deficient *Aspergillus* strains, e.g. by cloning a selectable
 CC marker into the CPY gene, provides suitable host strains for prodn.
 CC of heterologous protein.
 XX
 XX Sequence 556 AA;
 SQ

Alignment Scores:
 Pred. No.: 1,95e-48 Length: 556
 Score: 630.00 Matches: 154
 Percent Similarity: 48.45% Conservative: 81
 Best Local Similarity: 31.75% Mismatches: 174
 Query Match: 21.69% Indels: 76
 DB: 17 Gaps: 17

US-09-712-338-1_COPY_55_1662 (1-1608) x AAR96737 (1-556)

QY	79	GCAAAACATGTCACCATCGGTACAGGAACCGGGGAGAGGGCGTCTCGAGAGTACC	138
Db	131	AlaTyrAspLeuArgValLysLysThrAspProGlySerLeuGlyLeu-----Asp	147
QY	139	CGGGGTGTCAAATCTACTCTGGATATGTCGACACCTCTCCC---GAGTCCCATACATTC	195
Db	148	ProGlyValLysGlnThrGlyTyrLeuAspAsnGluAsnAspLysHisLeuPhe	167
QY	196	TTCTGTTCTCTGGAACCGACATACACCACTGACCACTGATCATCATTTGGTTGAAT	255
Db	168	TyrTrpPhePheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn	187
QY	256	GGTGGCCCTGGAAGCGATTTCTTGATCGCTCTTCGAAGAGTGGGCCCTTGGCCATGIC	315
Db	188	GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle	207
QY	316	AATTCAGCTTTGATGACTACATCAACCTCTACTCTGTGAACGAGCTCCCATTTACTA	375
Db	208	AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle	227
QY	376	TTCTCTCTCCAGCATTTGGAGCGGTCTTCATAGTGATACAGGTGTGATGGTCCATT	435
Db	228	PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla-----	243
QY	436	AACCTGTGTAAGTGGGTGCTGCGAAATTCGAGCTTTGCAGGAGTTTCAGGGCGGTACCCA	495
Db	243	-----	243
QY	496	ACCATGATGCCACTGTATCGATACATCAACATCTGCGCAGAGCGCGTTCGGGATC	555
Db	244	-----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu	257
QY	556	CTGCAAGGATTCCTTAGTGCATACCTAGCTAGCTGAGGTGAGTCAAGGACTTC	615
Db	258	LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe	273
QY	616	AGTCTATGGACGAGAGCTATGGAGGCGCATATGTCCTGCATTCCTCAATCATTTTAC	675
Db	274	HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu	293
QY	676	GACGAGAATCAGAGATTTGCCAAGGTAGTGTANTGGTTCAGCTTAATTTCAACTCT	735
Db	294	SerHisLysLysArg-----AsnIleAsnLeuGlnSer	304
QY	736	CGGGGAATTAATACGGCATCATCGACGAGCGCATCCAGCCCTTACTACCTCGAATTC	795
Db	305	ValLeuIleGlyAsnGlyLeuThrAspGlyTyrThrGlnTyrGluTyrTyrArgProMet	324
QY	796	GCTGTGAACAATACCTACGGTATACAGGCTGTCAAGAGACCGGTCTACAACATCATGAAG	855

Db	325	AlaCysGlyAspGlyGlyTyrProAlaValLeuAspGluSerSerCysGlnSerMetAsp	344
QY	856	TTTGCCACCAATGCAATGGTTGCCAGGATTTCATTCCACCTGCAACAGACAAC	915
Db	345	-----AsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu	361
QY	916	CGCACCGCATTAGCTACGCTCTGCGCGGAAGCCACCAACATGTGCAGGGACAAT	975
Db	362	SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla	376
QY	976	GTGAGGGGCCATATACGCTTGTGCTGCTGGTGTGTATGATCAATTCGGCATCATAT	1035
Db	377	LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys	395
QY	1036	GATGACCCG-----ACTCCGCCAAGTTATTACAAACAAATTTCTGCGCAAG	1080
Db	396	GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys	415
QY	1081	GACTCTGTGATGAGCGCTATCGGCGTCAACATCAAC---TACACCCAGTCCCAATATGAC	1137
Db	416	ProGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp	435
QY	1138	GTCCTACTACGCTTCCAGCAACAGCGGACTTTGTCTGGCCC-----AACITTCATCGAA	1191
Db	436	IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro	455
QY	1192	GACCTCGAGGAGATCTTGTCTCCCGTGTCTCCCTCATCAT---GGCGAGGCC	1248
Db	456	GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla	470
QY	1249	GATTACATCTCAACTGGTTCGGCGGTGAGCGGCTTCCCTCGCTCGGAACCTACTCCCA	1308
Db	471	AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly	490
QY	1309	GCGGCCGATCTCCGAGCGGAGGTACACGCCCTGTGAAATC-----AAC	1353
Db	491	GlnAlaGluTyrAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr	510
QY	1354	GGCGTCGATGATGGGAAACTCGGAGTATGTAATTTCTCTCACTCGCTCTATGAG	1413
Db	511	GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly	530
QY	1414	GCAGGCATGAATGCCATACCTACAGCCCATGCCCTCCCTGCAATTTGTTAACCGACT	1473
Db	531	GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp	550
QY	1474	ATCTTCGGT---TGG	1485
Db	551	LeuGlyGlyGluTrp	555

RESULT 9
 AAR96738
 ID AAR96738 standard; Protein; 557 AA.
 XX
 AC AAR96738;
 XX
 XX 14-AUG-1996 (first entry)
 DT
 XX
 XX A. niger SFAG 2 carboxypeptidase Y.
 DE
 XX Carboxypeptidase Y; CPY; ascomycete; deuteromycete; host cell;
 KW protease deficiency.
 KW
 XX Aspergillus niger strain SFAG 2.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..138
 FT /label= Pre-pro_peptide
 FT Protein 139..557
 FX /label= Mat_protein
 XX
 PN W09609397-A1.

28-MAR-1996.
 19-SEP-1995; 95WO-US11945.
 20-SEP-1994; 94US-0309341.
 (NOVO) NOVO NORDISK BIOTECH INC.
 Thompson SA, Yaver DS;
 WPI; 1996-188458/19.
 N-PSDB; AAT28284.
 Nucleic acid construct encoding a filamentous ascomycete or deuteromycete carboxypeptidase Y - useful to produce host cells modified to produce reduced amounts of carboxypeptidase
 Claim 5; Page 23-25; 46pp; English.
 Carboxypeptidase Y (AAR96738), a vacuolar protease, is the product of the CPY gene (AAT28284) of *Aspergillus niger* strain SFAG 2. Creation of CPY-deficient *Aspergillus* strains, e.g. by cloning a selectable marker into the CPY gene, provides suitable host strains for prodn. of heterologous protein.
 Sequence 557 AA;
 Alignment Scores:
 Pred. No.: 1.6e-47 Length: 557
 Score: 620.00 Matches: 154
 Percent Similarity: 48.25% Conservative: 80
 Best Local Similarity: 31.75% Mismatches: 175
 Query Match: 21.35% Indels: 76
 DB: 17 Gaps: 18
 US-09-712-338-1_COPY_55_1662 (1-1608) x AAR96738 (1-557)
 QY 79 GCAACAATGTCACCATCGGTACAGGAACCGGCGAGGCGGTGCGAGACTACC 138
 DB 132 AlATyRAspLeuGValLysLysThrAspProSerSerLeuGlyIle-----Asp 148
 QY 139 CCGGGTGTCAAACTCTACTCTGATATGTCACACCTCTCC-----GAGTCCCATACCTTC 195
 DB 149 ProGlyValLysGlnTyThrGlyTyLeuAspAspAsnGluAsnAspLysHisLeuPhe 168
 QY 196 TTCTGGTCTTCGAGGCCACATACCCAGAACTGACATACATGTTGTTGAAT 255
 DB 169 TyrIrrpPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
 QY 256 GTGGCCCTGGAAGCGATTCTTTCATCGGTCTCTCGAAGAGTTGGCCCTTGCCATGTC 315
 DB 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
 QY 316 AATTCGACTTTTGAATGACTACATCAACCTCATCTGCTGGAACGAGGTCTCCAAFTTACTA 375
 DB 209 AsnLysLysIleGlnProValTyAsnAspTyAlaIrrpAsnSerAsnAlaSerValle 228
 QY 376 TTCTGTGTCGAGCATTTGGAGTCGGCTTTTCATATAGTATAGTACGGTTGATGGTCAAT 435
 DB 229 PheLeuAspGlnProValAsnValGlyTySerTyrSerAsnSerAla----- 244
 QY 436 AACCTGTAACTGGGGTCTGCGAAATTCGAGGCTTCGAGGAGTTCAGGCGCGGTACCCA 495
 DB 244 ----- 244
 QY 496 ACCATTGATGCCACTCTGATCGATACATACCANTTTGCCGAGAGCGCGTGTGGAGATC 555
 DB 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyAlaLeu 258
 QY 556 CTCGAGGATTCCTTAGTGACTACCTAGCTTGGACTCTGAGTGTAGGAGTTC 615
 DB 259 LeuThrLeuPheLysGlnPhePro-----GluTyAlaLysGlnAspPhe 274

QY 616 AGTCTATGACGGAGAGCTATGAGGGCAGCTATGTCCTCGCATTTCTCAATCATTTTAC 675
 DB 275 HisIleAlaGlyGluSerTyAlaGlyHisTyIleProValPheAlaSerGluIleLeu 294
 QY 676 GAGCAGAATGAGAGAAATGCCAACGCTAGTGTATGTTGTTGTTGTTGTTGTTGTTGTT 735
 DB 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
 QY 736 CTGGGAATTAATTACGGCATCATCAGCAGGCGATCCAGGCCCTTACCTCACTCAATTC 795
 DB 306 ValLeuIleGlyAsnGlyLeuThrAspGlyLeuThrGluTyGluTyArgProMet 325
 QY 796 GCTGTGAACATACCTACGCTATCAAGGCTGTCAAGCAGACGCTGTCAACTACTATGAG 855
 DB 326 AlaCysGlyAspGly---GlyTyProAlaValLeuAspGlu---GlySerCysGlnAla 343
 QY 856 TTTCGCAACCAATGCCAAATGGTGGCCAGGATTTGATTTCCACCTGCAACAGACAAC 915
 DB 344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTySerSerGlu 362
 QY 916 CGACCCGATTAAGTACTACGCTCTGCGCCGAGCCACACATGTCGAGGAGCAAT 975
 DB 363 SerAla-----TrpValCysValProAlaSerIleTyCysAsnAla 377
 QY 976 GTTCAGGGGCCATACTACGCTTCTGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1035
 DB 378 LeuLeuAlaProTyArgGlnArg---ThrGlyGlnAsnValTyAspValArgGlyLysCys 396
 QY 1036 GATGACCGG-----ACTCGCCCAAGTTATTACACAAATTTCTGCAAG 1080
 DB 397 GluAspSerSerAsnLeuCysTySerAlaMetGlyTyValSerAspTyLeuAsnLys 416
 QY 1081 GACTCTGTGATGAGCGCTATCGGCTCAACATCAAC---TACACCCAGCTCCCAATATGAC 1137
 DB 417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyAspSerCysAsnPheAsp 436
 QY 1138 GTCTACTACGCTTTCCAGCAACAGGCGACTTTGCTGGCC-----AATTCAICAA 1191
 DB 437 IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyHisArgLeuValPro 456
 QY 1192 GACCTCGAGGAGACTCTTGTCTCCCGTGTCTCCCTCACTAT---GGCGACGCC 1248
 DB 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyAlaGlyAspAla 471
 QY 1249 GATTACATCTCAACTGTTGCGCGGTCAAGCCGTTTCCCTCGCTGCGAAGTACTCCCAA 1308
 DB 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
 QY 1309 GCCGCCAGTTCGAGGCGAGGTACAGCCCTGAAAGTC-----AAC 1353
 DB 492 GlnAlaGluTyAlaSerAlaLysLeuGluAspLeuValValGluAsnGluHisLys 511
 QY 1354 GCGCTCGAGTATGGGAAACTCGCGAGTATGTAATTTCTCTCTCCCTCACTGCTATGAG 1413
 DB 512 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyGly 531
 QY 1414 GCAGGCCATGAAGTCCCATACTACAGCCCATCCCTCCCTGCAATTTGTTTAAACGGACT 1473
 DB 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 551
 QY 1474 ATCTTCGGT---TGG 1485
 DB 552 LeuGlyGlyGluTrp 556
 RESULT 10
 AAR48059
 ID AAR48059 standard; Protein: 491 AA.
 XX
 AC AAR48059;
 XX
 DT 25-MAR-2003 (updated)
 DT 20-JUL-1994 (first entry)

XX Sequence of protease C encoded by gene K1.PRC1.
 DE
 XX Protease; yeast; proteolysis.
 KW
 XX Kluyveromyces lactis.
 OS
 XX WO9400579-A1.
 FN
 XX 06-JAN-1994.
 PD
 XX 23-JUN-1993; 93WO-FR00623.
 XX
 XX 25-JUN-1992; 92FR-0007785.
 XX
 XX (RHON) RHONE POULENC RORER SA.
 PA
 XX Fleer R, Fournier A, Yeh P;
 FI
 XX WPI; 1994-026215/03.
 DR
 XX N-PSDB; AAQ55347.
 DR
 XX New Kluyveromyces yeast with modified protease gene - esp. used
 PT for high yield prodn. of recombinant protein, also DNA encoding
 PT yeast protease and derived peptide(s)
 PT
 XX Disclosure; Page 28-31; 49pp; English.
 PS
 XX The protease gene is to be modified in order to render it (partially)
 CC incapable of producing the natural protein; or result in a non-
 CC functional protease or in a protease with modified proteolytic
 CC activity. The modifications can be introduced in vitro or in situ by
 CC standard genetic engineering techniques or by exposure to mutagenic
 CC agents.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX Sequence 491 AA;
 SQ

Alignment Scores:
 Pred. No.: 7.52e-43 Length: 491
 Score: 568.50 Matches: 151
 Percent Similarity: 44.65% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 184
 Query Match: 19.58% Indels: 85
 DB: 15 Gaps: 14

US-09-712-338-1_COPY_55_1662 (1-1608) x AAR48059 (1-491)

Qy 79 GCAACAATGTCACATCCCGTACAGAGAACCCGGGCGAGGCGGTGTCGAGACTACC 138
 Db 62 AlatyrsSerLeuArgLleLysProLeuAspProLysSerLeuGlyVal----- 77
 Qy 139 CCGGGTGCRAATCCTACTCTGGATATGTCACACCTCTCCGAGTCCCATACCTCTTC 198
 Db 78 AspThrValLysGlnTrpSerGlyTyrLeuAspTyrGlnAspSerLysHisPhePheTyr 97
 Qy 199 TGGTTCCTCGAAGCCAGCATACACCAAACTGCACATATCATCATCTGGTTGAATGCT 258
 Db 98 TrpPheGluSerArgAsnAspProGluAsnAspProValLleLeuTrpLeuAsnGly 117
 Qy 259 GGCCCTCGAAGCGATCTCTTGATCGGTCTCTTCGAGAGAGTGGGCGCTGCCATGTCAAT 318
 Db 118 GlyProGlycysSerSerPheValGlyLeuPheGluLeuGlyProSerSerileGly 137
 Qy 319 TCGACTTTTGATGACTACATCAACCTCAGTCGTGGACGAGGCTCCCAATTTACTATTC 378
 Db 138 AlaAspLeuLysProLleTyrAsnProTyrSerTrpAsnSerAsnAlaSerValilePhe 157
 Qy 379 CTGTCGCCAGCATTTGGAGTGGGCTTTTCATATAGTATGATCGGTGTGTCATTAAC 438
 Db 158 LeuAspGlnProValGlyValGlyPheSerTyrGlyAsp----- 170
 Qy 439 CCTTAAGTGGGGTCGCGAAATTCGAGCTTTCGAGGAGTTCAGGGCGCGTACCCAACC 498

Db 170 ----- 170
 Qy 499 ATTGATGCCACTCTGATCGATACCAATCTTCCGCGAGAGCGGCTTGGAGACTCGT 558
 Db 171 -----SerLysValSerThrAspAlaLysAspValTrilePheLeu 187
 Qy 559 CAGGATTCCCTTAGTGGACTACCTAGCTTAGGTGCGAGTCTAGGACTTCACTAGT 618
 Db 188 AspLeuPheGluArgPheProHisLeu-----ArgAsnAsnAspPheHis 203
 Qy 619 CTATGACGGAGAGCTATGAGGCGACTATGCTCTGCATCTTCAATCATCTTTACGAG 678
 Db 204 IleSerGlyGluSerTyrAlaGlyHisTyrLeuProLysLleAlaHis----- 219
 Qy 679 CAGAATGAGAGAAATGCCAACGGTAGTCTTAATGGTGTTCAGCTTAATTTCACTCTCT 738
 Db 220 -----GluIleAlaValHisAlaGluAspSerSerPheAsnLeuSerVal 236
 Qy 739 GGAATTAATAACGGCATCATCGACGAGCGCATCCAGGCCCTTACTACCTGAAATCGCT 798
 Db 237 LeuIleGlyAsnGlyPheThrAspProLeuThrGlnTyrGlnTyrGluProMetAla 256
 Qy 799 GTGAACAATACCTACGGTATCAAGCTGTCAACGAGACCGCTCTACACTACATCAAGTTT 858
 Db 257 CysGlyGluGly--GlyTyrProAlaValLeuGlu----- 267
 Qy 859 GCCAACCAAAATGCCAAATGGTTCAGGAT----- 888
 Db 268 -----ProGluAspCysLeuAspMetAsnArgAsnLeuProLeuCysLeuSer 283
 Qy 889 TTGATTTCCACTGCAACAGACAAACCGCACCGCATTTAGCTAGCTACGCCCTCTGCC 948
 Db 284 LeuValAspArgCysTyrLysSerHis-----SerValPheSerCysVal 298
 Qy 949 GAAGCCACCAACATGTGACGGGACAAATGTTGAGGGGCCATCTACTACGCTTTCGCTGGT 1008
 Db 299 LeuAlaAspArgTyrCysGluGlnGlnIleThrGly---ValTyrGluLysSerGlyArg 317
 Qy 1009 GGTGTGTATGATATTCGG-----CATCCATATGATGACCG-----ACT 1047
 Db 318 AsnProTyrAspIleArgSerLysCysGluAlaGluAspAspSerSerGlyAlaCysTyrGln 337
 Qy 1048 CCGCCAAAGTTATTACAACAAAATTTCTGCCAAGGACTCTGTCAAGAGCTATCGCGCTC 1107
 Db 338 GluGluIleTyrIleSerAspTyrLeuAsnGlnGluValGlnArgAlaLeuGlyThr 357
 Qy 1108 AACATCACTACACCCAGTCC---AATAATGACGTCTACTAGCTTTCCAGCAACAGGC 1164
 Db 358 AspValSerSerPheGlnGlyCysSerSerAspValGlyIleGlyPheAlaPheThrGly 377
 Qy 1165 GACTTTGCTGGCCCAACTTTCATCGAGACCTCGAGAGATCCTTTCCTCCCGCTGGT 1224
 Db 378 AspGlyProSerPro---PheHisGlnTyrValAlaGluLeuLeuAspGlnAspIleAsn 396
 Qy 1225 GTCCTCCCTACTATGCGGACGCCCATATCATCTGCACACTGCTTCGGGCGGTCAGCCGTT 1284
 Db 397 ValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGlyAsnLeuAlaTrp 416
 Qy 1285 TCCCTCGCTCGGAACACTACTCCCAAGCCCGCCAGTTCGGAAGCGCAGGTACAGCCCGCTG 1344
 Db 417 ThrGluLysLeuGluTrpArgTyrAsnGluGluTyrLysLysGlnValLeuArgThrTrp 436
 Qy 1345 AAAGTCAACGGCGTGAG-----TATGGGAAACTCGCGATATGTAATTCPCCTTC 1398
 Db 437 LysSerGluGluThrAspGluThrIleGlyGluThrLysSerTyrGlyProLeuThrTyr 456
 Qy 1399 ACTCCGCTCTGAGGAGGCGCATCAAGTCCCATCTACTACCCCATCGCTCCCTCGCAA 1458
 Db 457 LeuArgIleTyrAspAlaGlyHisMetValProHisAspGlnProGluAsnSerLeuGln 476
 Qy 1459 TTGTTTAACCGGACTATC 1476
 Db 1459 ----- 1476

Db 477 MetValAsnSerTrpIle 482

RESULT 11

ID ABG93281

XX ABG93281 standard; Protein; 550 AA.

AC ABG93281;

XX 21-NOV-2002 (first entry)

XX C. albicans BAX-associated protein fragment SEQ ID 520.

DE

XX Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;

KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;

KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;

KW neurodegeneration; cell death.

XX Candida albicans.

OS

XX WO200264766-A2.

XX 22-AUG-2002.

XX 21-DEC-2001: 2001WO-EP15398.

XX 22-DEC-2000: 2000EP-0870318.

PR 04-JAN-2001: 2001EP-0870002.

PR 09-JAN-2001: 2001EP-0870003.

XX (JANC) JANSSEN PHARM NV.

PA

XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

PI WPI: 2002-667002/71.

DR N-PSDB; ABQ76547.

XX

XX New isolated nucleic acid representing a synthetic Bax gene, useful as

PT medicament for treating, preventing and/or alleviating yeast or fungal

PT infections or proliferative disorders, or for preventing apoptosis in

PT certain diseases

XX

PS Claim 36; Figure 2: 344pp; English.

XX

CC This invention describes a novel nucleic acid representing a synthetic

CC Bax gene. The Bax gene of the invention is useful for identifying

CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying

CC Candida spp. sequences that are differentially expressed in a pathway

CC eventually leading to programmed cell death or identifying inhibitors or

CC inhibitor sequences of Bax-induced cell death. The products of the

CC invention have cytostatic, fungicide; immunosuppressive, virucide and

CC vasotropic activity and can be used in vaccines or for gene therapy. The

CC isolated nucleic acids, polypeptides, pharmaceutical compositions,

CC antisense molecules and antibodies are useful as medicaments or in

CC preparing a medicament for treating, preventing and/or alleviating

CC diseases associated with yeast or fungi or proliferative disorders, such

CC as cancer, or for preventing apoptosis in certain diseases. The compounds

CC or polypeptides, or the genetically modified organism are useful for

CC preparing a medicament for modifying the endogenous flora of humans and

CC other mammals. The vaccine is useful for immunising against yeast or

CC fungal infections. Apoptosis-related diseases include autoimmune disease,

CC ischaemia, diseases related with viral infections or neurodegenerations.

CC This sequence represents a polypeptide associated with the Bax gene

CC described in the disclosure of the invention.

XX

SQ Sequence 550 AA;

Alignment Scores:

Pred. No.:	2,31e-40	Length:	550
Score:	541.50	Matches:	150
Percent Similarity:	44.47%	Conservative:	63
Best Local Similarity:	31.32%	Mismatches:	157
Query Match:	18.65%	Indels:	109
DB:	23	Gaps:	17

US-09-712-338-1_COPY_55_1662 (1-1608) x ABG93281 (1-550)

QY 145 GTCAATCTCTACTCTGGATATGTCGAC---ACCTCTCCGAGTCCCATACCTCTCTCTGG 201

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 133 VailysGlnTyrThrGlyTyrLeuAspIleAspSerLeuAspLysHisLeuPheTyrTrp 152

QY 202 TTCTTCGAAGCCACACATACCCAGAAACTGCACCTATCACATTTGGTTGATGTTGGTGGC 261

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 153 PhePheGluSerArgAsnAspProLysAsnAspProIleIleLeuTyrLeuAsnGlyGly 172

QY 262 CCTGGAAGCGATCTTTGATCGGTCTCTGGAAGAGTTGGCCCTTGCATCTGCAATTCG 321

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 173 ProGlyCysSerSerSerThrGlyLeuPhePheGluLeuGlyProSerSerIleAsnLys 192

QY 322 ACTTTTGATGACTACATCAACCCCTCACTCGTGGAGAGGCTCTCAATTACTTACTCTG 381

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 193 ThrLeuHisProValTyrAsnProTyrSerIleAsnSerAsnAlaSerValIlePheLeu 212

QY 382 TCCAGGCATGGAGTCGGCTTTTCATATAGTCAIACGGTTGATGGTTCATTAACCCCT 441

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 213 AspGlnProValGlyValGlyTyrSerTyr----- 222

QY 442 GTAACGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 501

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 223 ---ThrGlyGlyAspGlu----- 227

QY 502 GATGCCACTCTGATPCGATACCTACCAATCTTCCGAGAGGCGGCTTGGAGATCTGCAA 561

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 228 -----ValLysAsnThrLeuThrAlaAlaLysAspValTyrValPheLeuGlu 243

QY 562 GGATTCCTTAGTGGACTACCTAGCTGGACTCTAGGCTGAGGTGAGTCTAGGCTTCACTCA 621

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 244 LeuPhePheGlnLysPhePro-----GlnPheLeuThrAsnLysPheHisIle 259

QY 622 TGGACGAGAGCTATGGAGGCGACTATGGTCCCTCATCTTCAATCATTTTACGAGCAG 681

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 260 AlaGlyGlnSerTyrAlaGlyHisTyrIleProAlaPhe----- 272

QY 682 AATGAGAGAATGCCAACGGTGTATGTTAATGGTTCAGCTTAATTTCACTCTCTCTGGGA 741

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 273 -----AlaSerGluIleIleAsnAsnAlaAspArgSerPheGluLeuAlaSer 288

QY 742 ATTATT-----AACGGCATCTCCAGAGCGATCCAGGCCCTTACTACCTGGAATTC 795

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 289 ValLeuIleGlyAsnGlyIleThrAspProLeuIleGlnAspGlySerTyr----- 305

QY 796 GCTGTGAACAATACCTACCGTATCAGGCTGTCAACGAGACCGCTCTACAATCATCATGAG 855

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 305 ----- 305

QY 856 TTTGCCAACCAATGCCAAATGGTTC-----CAGGATTTGATTTCACCC 900

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 306 -----LysProMetGlyCysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 321

QY 901 -----TGCAACAGACAAACCCACCCATCTAGCTACCTACCGCCCTCTGGCGCAGACC 954

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 322 GlnGlnCysAspGlnMetGluArg-----AspTyrProArgCysAlaLysLeu 337

QY 955 ACCAACATGTGTCAGGAC-----AATGTGTGAGGGGGCCATAC----- 990

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 338 ThrLysLeuCysTyrSerPheGlnSerAlaLeuThrCysValProAlaGlnTyrTyrCys 357

QY 991 -----TACGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1029

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 358 AspSerArgLeuPheGlnProTyrAlaGlnThrGlyLeuAsnProTyrAspIleArgLys 377

QY 1030 CCATATCATGACCCG-----ACTCCGCCAAGTATTACAAACAATTTCTG 1074

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 378 AspCysAlaGluGlnGlyAsnCysTyrValGluMetAspTyrLeuAspGluTyrLeu 397

QY 1075 GCAAAGGACTCTCTCATGAGCGCTATCGGGCTC---AACATCAAC---ATACCCAGTCC 1128

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
Db      398 AsnLeuAspTyrVallysGluAlaValGlyAlaSerAsnIleAspIlePheThrSerCys 417
QY      1129 AATAATGACGCTACTACGCTTCCAGCAACAGCGAGCTTGTCTGGCCCAACTTCATC 1188
Db      418 AspAspThrValPheArgAsnPhelIleLeuAspGlyAspGluMetIysPro---PheGln 436
QY      1189 GAAGACCTGGAGGAGCTCTTGTCTCCCGTGGTGTCTCCCTCATCTATGCGAGGCC 1248
Db      437 GlnTyrValAlaGluLeuLeuAspAsnValProValLeuIleTyrAlaGlyAspLys 456
QY      1249 GATTACATCGCAACTGGTTCGGCGTCAGCGGTTCCCTCGCTCGGAACACTACTCCAA 1308
Db      457 AspTyrIleCysAsnTrpLeuGlyAsnLeuAlaTrpValAsnGluLeuGluTyrSerAsp 476
QY      1309 GCGGCCAGTTCGGAAGCGCAGGTACAGCCCTGAAAGTCAACGCGCTCGAG--TAT 1365
Db      477 SerGluHisPheAlaProLysProLeuGlnLeuTrpLysGlnAspGlyLysLysAlaAla 496
QY      1366 GGGGAACACTCGCGAGTATGTAATTCTCTTCACCTCGCGTCATGAGCGAGCCATGAA 1425
Db      497 GlyGluValLysAsnHisLysHisPheThrPheLeuArgIleTyrAspAlaGlyHisMet 516
QY      1426 GTCCCATACTACAGCCCTCGCTCCCTGCAATTTGTTAAACCGGACTATCTTCGGT 1482
Db      517 ValProPheAspGlnProGluAsnAlaLeuSerMetValAsnThrTrpValGlnGly 535

RESULT 12
ABR38845
ID   ABR38845 standard; Protein; 481 AA.
XX
AC   ABR38845;
XX
DT   24-APR-2003 (first entry)
XX
DE   A. niger carboxypeptidase Y polypeptide #3.
XX
KW   Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW   protein solubility; viscosity; taste; texture; nutritional value;
KW   EC3.4.16.5.
XX
OS   Aspergillus niger.
XX
PN   WO200268623-A2.
XX
PD   06-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-EP01984.
XX
PR   23-FEB-2001; 2001EP-0200657.
PR   23-FEB-2001; 2001EP-0200658.
PR   23-FEB-2001; 2001EP-0200660.
PR   26-FEB-2001; 2001EP-0200706.
PR   26-FEB-2001; 2001EP-0200707.
PR   26-FEB-2001; 2001EP-0200708.
PR   26-FEB-2001; 2001EP-0200719.
PR   28-MAR-2001; 2001EP-0000075.
PR   28-MAR-2001; 2001EP-0000078.
PR   28-MAR-2001; 2001EP-0000080.
PR   28-MAR-2001; 2001EP-0000087.
PR   28-MAR-2001; 2001EP-0000088.
PR   21-MAY-2001; 2001EP-0000156.
PR   21-MAY-2001; 2001EP-0000159.
PR   21-MAY-2001; 2001EP-0000160.
PR   21-MAY-2001; 2001EP-0000162.
PR   21-MAY-2001; 2001EP-0000165.
PR   21-MAY-2001; 2001EP-0000166.
PR   21-MAY-2001; 2001EP-0000168.
PR   21-JUN-2001; 2001EP-0000240.
PR   21-JUN-2001; 2001EP-0000242.
PR   21-JUN-2001; 2001EP-0000244.
PR   21-JUN-2001; 2001EP-0000246.
PR   12-JUL-2001; 2001EP-0000280.
PR   12-JUL-2001; 2001EP-0000285.

PR      30-JUL-2001; 2001EP-0000323.
PR      30-JUL-2001; 2001EP-0000327.
PR      02-AUG-2001; 2001EP-0000341.
PR      02-AUG-2001; 2001EP-0000342.
PR      02-AUG-2001; 2001EP-0000343.
PR      02-AUG-2001; 2001EP-0000344.
PR      09-AUG-2001; 2001EP-0000357.
PR      16-AUG-2001; 2001EP-0000374.
PR      16-AUG-2001; 2001EP-0000377.
PR      20-SEP-2001; 2001EP-0000483.
PR      22-OCT-2001; 2001EP-0000552.
PR      22-OCT-2001; 2001EP-0000553.
PR      22-OCT-2001; 2001EP-0000554.
PR      22-OCT-2001; 2001EP-0000556.
PR      22-OCT-2001; 2001EP-0000557.
PR      22-OCT-2001; 2001EP-0000558.
PR      13-NOV-2001; 2001EP-0004464.
PR      21-DEC-2001; 2001EP-0005117.
XX
XX      (STAM ) DSM NV.
XX
PI      Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI      Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI      Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI      Stiebler J, Albang R;
XX
XX      WPI: 2002-723203/78.
DR      P-PSDB; ABZ78212, ABZ78269.
XX
XX      Novel isolated protease polypeptide useful in laboratory, clinical,
PT      pharmaceutical, chemical, diagnostic, personal care and industrial
PT      applications -
XX
XX      Claim 13; Page 335-337; 394pp; English.
XX
XX      The invention relates to a novel isolated protease polypeptide. A
CC      polypeptide or polynucleotide of the invention is useful for diagnosing a
CC      fungal infection such as aspergillosis, or as a query sequence to perform
CC      a search against public databases. A polypeptide of the invention is
CC      useful in a selected number of industrial or pharmaceutical processes, in
CC      laboratory or clinical processes, in food industry (baking, brewing, in
CC      cheese manufacturing, meat tenderising), in tanning industry and in the
CC      manufacture of biological detergents. A polypeptide may also be useful
CC      for improving protein solubility, extraction yields, viscosity or taste,
CC      texture, nutritional value, minimising of antigenicity or
CC      anti-nutritional factors, colour or functionality as well as processing
CC      aspects like filterability of the proteinaceous raw material. The
CC      sequences shown in ABR38313-ABR3869 represent the A. niger proteases of
CC      the invention.
XX
XX      Sequence 481 AA;
SQ
XX
XX      Alignment Scores:
XX      Pred. No.: 9.35e-29 Length: 481
XX      Score: 414.00 Matches: 140
XX      Percent Similarity: 41.73% Conservative: 77
XX      Best Local Similarity: 26.92% Mismatches: 207
XX      Query Match: 14.26% Indels: 96
XX      DB: 23 Gaps: 20
XX
XX      US-09-712-338-1_COPY_55_1662 (1-1608) x ABR38845 (1-481)
QY      43 CCCAAGAACCCACCGCGGGTCAAGACTCTT-----ACAACGCAACAATGTC 90
Db      23 ProGluGluProSerAspPheArgThrPheHisSerProTyrSerProHisHisSerIle 42
QY      91 ACCATCGGTTACAGGACACCGCGGCGAGCGGCTCTGCGAGACTACCCCGGGTGTCAA 150
Db      43 ArgIleArgGlnGlnAsn-----GluSerIleCysAlaAlaHisSerAla----- 57
QY      151 TCCTACTCTGATATGTGACACACCTCTCCCGAGTCCCATACCTTCTTGGTTCGTGAA 210
```


PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160983.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26 OCT-1999; 99US-0161361.

Db	270	AsnTyrAsnLysPhe	-----CysThrAsp-----SerAspLeu	280
Qy	904	AAACAGACAACCCGACCGCATAGCTAGCTACGACCGCTCTCGCGGAGCCACCAACATG	963	
Db	281	TyrAspIrpAspLysCysHisLeuAlaSerGlnLysIleGluAlaGlnLysThrHisLeu	300	
Qy	964	TGCAGGACAAATGTTGAGGGGCCATACTACGCCCTTGCTGCTGCTGCTGATGATAT	1023	
Db	301	-----AspIleTyrAsnIle	305	
Qy	1024	CGGCATCCATAT-----	1035	
Db	306	TyrAlaIrpLeuCysLeuAsnSerThrLeuSerSerGluProLysLysCysThrThrIle	325	
Qy	1036	-----GATGACCGGACTCCGCCCAAGTTATTACAAATAATCTGGCAAGGACTCGTC	1089	
Db	326	MetIysAlaAspProCysSerSerGlyAsnTyrLeuLysAlaTyrLeuAsnIleLysGluVal	345	
Qy	1090	ATGGACGCTATPCGCGCTCAAC-----ATCAACTACACCCAGTCCCAATAT	1134	
Db	346	GlnGluAlaIleHisAlaAsnThrThrLysIleProTyrGluIrpThrSerCysAsnThr	365	
Qy	1135	GACGTCTACTACGTTTCCAGCAACAGGCGACTTCTGTCGGCCCACTTCATCGAAGAC	1194	
Db	366	LysLeuLeuIrpGluIrpAsnGluLysAspArgTyrVal-----SerLeuIrpProIle	383	
Qy	1195	CTCGAGGAGATPCCTTGCTCTCCCGTCGCGTCTCCCTCATCTATCGGACGCGGATTAC	1254	
Db	384	GlnGlnGluLeuMetGlyLysGlyValArgValMetLeuTyrAsnGlyAspValAspLeu	403	
Qy	1255	ATCTGCAACTGGTTGGCGGTGTCAGCGGTTCCCTCGCTCGCAACTACTCCCAAGCGCC	1314	
Db	404	ValIleProPheThrSerThrLeuAlaValValLysThrMetAsnLeuThrValValLys	423	
Qy	1315	CAGTTCGAGCGCAGGATACAGCCCTCAAAGTCAACGCGGTGAGTATGGGAACT	1374	
Db	424	GluIrpArg-----ProTrpPheThrGlyHisValGlyGlyPheThr	438	
Qy	1375	CGCGAGTAT--GGTAATTTCTCCTTCACTCGCTCTATGAGCGAGCGCATGAAGTCCCA	1431	
Db	439	GluAspTyrLysGlyAsnLeuThrPheValThrValLysGlyAlaGlyHisSerValPro	458	
Qy	1432	TACTACAGCCCATCGCTCCCTGCATATGTTTAAACGGCATATC	1476	
Db	459	ThrAspGlnProIleHisAlaLeuAsnIlePheThrSerPheIle	473	
RESULT 14				
XX	AAG30064			
XX	ID	AAG30064 standard; Protein; 502 AA.		
XX	AC	AAG30064;		
XX	XX			
XX	DE	17-OCT-2000 (first entry)		
XX	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 35876.		
KW	DE	Protein identification; signal transduction pathway; metabolic pathway;		
KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	KW	termination sequence.		
OS	XX	Arabidopsis thaliana.		
XX	OS			
PN	PN	EP1033405-A2.		
XX	XX			
PD	PD	06-SEP-2000.		
XX	XX			
XX	XX	25-FEB-2000; 2000EP-0301439.		
PR	PR	25-FEB-1999; 99US-0121825.		
PR	PR	05-MAR-1999; 99US-0123180.		
PR	PR	09-MAR-1999; 99US-0123548.		
PR	PR	23-MAR-1999; 99US-0125788.		
PR	PR	25-MAR-1999; 99US-0126264.		

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144087.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

ID ABR38843 standard; Protein; 455 AA.

XX ABR38843;

XX 24-APR-2003 (first entry)

XX A. niger carboxypeptidase Y polypeptide #2.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;
KW EC3.4.16.5.

XX Aspergillus niger.

XX WO200268623-A2.

XX 06-SEP-2002.

XX 22-FEB-2002; 2002WO-EP01984.

XX 23-FEB-2001; 2001EP-0200657.

XX 23-FEB-2001; 2001EP-0200658.

XX 23-FEB-2001; 2001EP-0200660.

XX 26-FEB-2001; 2001EP-0200706.

XX 26-FEB-2001; 2001EP-0200707.

XX 26-FEB-2001; 2001EP-0200708.

XX 26-FEB-2001; 2001EP-0200719.

XX 28-MAR-2001; 2001EP-0000075.

XX 28-MAR-2001; 2001EP-0000078.

XX 28-MAR-2001; 2001EP-0000080.

XX 28-MAR-2001; 2001EP-0000087.

XX 28-MAR-2001; 2001EP-0000088.

XX 21-MAY-2001; 2001EP-0000156.

XX 21-MAY-2001; 2001EP-0000159.

XX 21-MAY-2001; 2001EP-0000160.

XX 21-MAY-2001; 2001EP-0000162.

XX 21-MAY-2001; 2001EP-0000165.

XX 21-MAY-2001; 2001EP-0000166.

XX 21-MAY-2001; 2001EP-0000168.

XX 21-JUN-2001; 2001EP-0000240.

XX 21-JUN-2001; 2001EP-0000242.

XX 21-JUN-2001; 2001EP-0000244.

XX 21-JUN-2001; 2001EP-0000246.

XX 12-JUL-2001; 2001EP-0000280.

XX 12-JUL-2001; 2001EP-0000285.

XX 30-JUL-2001; 2001EP-0000323.

XX 30-JUL-2001; 2001EP-0000327.

XX 02-AUG-2001; 2001EP-0000341.

XX 02-AUG-2001; 2001EP-0000342.

XX 02-AUG-2001; 2001EP-0000343.

XX 09-AUG-2001; 2001EP-0000357.

PI Klugbauer S, Wagner C, Fritz A, Albermann K, Stock A, Kimpel E;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albarg R;
XX WPI; 2002-723203/78.

XX (STAM) DSM NV.

XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, Albermann K, Stock A, Kimpel E;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albarg R;
XX WPI; 2002-723203/78.

DR P-PSDB; ABZ78210, ABZ78267.

XX Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications

XX Claim 13; Page 329-331; 394pp; English.

XX The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderizing), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABR38313-ABR38869 represent the A. niger proteases of
CC the invention.

XX Sequence 455 AA;

Alignment Scores:

Prod. No.: 1-97e-24 Length: 455
Score: 366.50 Matches: 138
Percent Similarity: 40.83% Conservative: 69
Best Local Similarity: 27.22% Mismatches: 182
Query Match: 12.62% Indels: 118
DB: 23 Gaps: 22

US-09-712-338-1_COPY_55_1662 (1-1608) x ABR38843 (1-455)

QY 7 GGAAGTACACGGCGTCCGGTACAGACACACTACCCAGAACCCACCCGGGGTCAAG 66
DB 25 GlyAlaAspProPheThrValPheArgSer-----ProHisSerProAla----- 39
QY 67 ACTCTTACACACCGCAACCAATGTCACCATCGGTACAGGACCCCGGGCAGAGGGCGTC 126
DB 40 -----PheSerIleArgIleGlnGlnAsn---AspSerIle 51
QY 127 TCGGAGACTACCCCGGGTGTCAAACTCTCTGTGATATGTCGACACCTCTCCGAGTCC 186
DB 52 CysAspAlaArgSer-----ProGlnPheThrGlyTrpLeuAspIleGlyProLys--- 68
QY 187 CATACCTTCTCTGGTTCCTCGAAGCCAGACATACCCAGAACTGCATATCATTCATG 246
DB 69 HisLeuPhePheTrpTyrPheGluSerGlnAsnAspProPheHisAspProLeuThrLeu 88
QY 247 TGGTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTCCGAAGATTGGCCCT 306
DB 89 TrpMetThrGlyGlyProGlyAspSerSerSerMetIleGlyLeuPheGluGluValGlyPro 108
QY 307 TCCCATGTCAAT-----TCGACTTTTGATGACTACATCAACCTCATCTCGTGGACGAG 360
DB 109 CysArgIleAsnGluPheGlyAsnGlyThrAspHis---AsnProTrpAlaTrpThrLys 127
QY 361 GTCTCCAATTACTATTCCTGTCGCCAGCCATTGGGAGTGGGCTTTTCATATAGTGATAG 420
DB 128 AsnSerSerLeuLeuPheValAspGlnProValAspValGlyPheSerTyrIleAsp--- 146
QY 421 GTTGATGGGTCCATTAAACCTGTAACTGGGGTCTGCGAAATTCGAGCTTTCAGGAGTT 480
DB 147 -----GluGlyTyr 149
QY 481 CAGGCCCGGTACCCAAACCATTTGATGCTGATGATGATGATGATGATGATGATGATGAT 540
DB 150 Glu-----LeuProHisAspSerArgGluAlaLaval 160
QY 541 GCCGCTTGGAGATCGTCGAGGATTCCTTAGTGACCTACCTAGCTAGCTAGCTAGCTAG 600
DB -----

Search completed: September 16, 2003, 18:41:58
Job time : 101.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:43:14 ; Search time 192 Seconds
(without alignments)
2493.964 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662

Perfect score: 2904

Sequence: 1 cttccaggaagtagcacggc.....gcattgtccagtgttgatg 1608

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 1112538

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09712338/runat_16092003.144327.14435/app_query.fasta.1.1799
-DB=published.Applications_AA_QFWT-fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCI=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09712338 @CGN.1.1.15 @runat.16092003.144327.14435
-NCFU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LANGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/BCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARTES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	517.5	17.8	421	9	US-09-420-785A-4	Sequence 4, Appli
2	517.5	17.8	421	9	US-09-901-252-15	Sequence 15, Appl
3	361	12.4	476	14	US-10-084-018-3	Sequence 3, Appli
4	359	12.4	476	10	US-09-909-320-164	Sequence 164, App
5	359	12.4	476	10	US-09-909-088B-164	Sequence 164, App
6	359	12.4	476	10	US-09-905-291A-164	Sequence 164, App
7	359	12.4	476	10	US-09-902-853-164	Sequence 164, App
8	359	12.4	476	10	US-09-907-841-164	Sequence 164, App
9	359	12.4	476	10	US-09-907-824-164	Sequence 164, App
10	359	12.4	476	11	US-09-904-011-164	Sequence 164, App
11	359	12.4	476	11	US-09-906-742-164	Sequence 164, App
12	359	12.4	476	11	US-09-906-838-164	Sequence 164, App
13	359	12.4	476	11	US-09-907-613-164	Sequence 164, App
14	359	12.4	476	11	US-09-907-942-164	Sequence 164, App
15	359	12.4	476	11	US-09-907-753-40	Sequence 40, Appl
16	359	12.4	476	11	US-09-904-859-164	Sequence 164, App
17	359	12.4	476	11	US-09-909-204-164	Sequence 164, App
18	359	12.4	476	11	US-09-904-820-164	Sequence 164, App
19	359	12.4	476	11	US-09-904-786-164	Sequence 164, App
20	359	12.4	476	11	US-09-906-646-164	Sequence 164, App
21	359	12.4	476	11	US-09-906-700-164	Sequence 164, App
22	359	12.4	476	11	US-09-903-786-164	Sequence 164, App
23	359	12.4	476	11	US-09-902-903-164	Sequence 164, App
24	359	12.4	476	11	US-09-903-749A-164	Sequence 164, App
25	359	12.4	476	11	US-09-904-119-164	Sequence 164, App
26	359	12.4	476	11	US-09-904-956-164	Sequence 164, App
27	359	12.4	476	11	US-09-902-736-164	Sequence 164, App
28	359	12.4	476	11	US-09-907-794-164	Sequence 164, App
29	359	12.4	476	11	US-09-903-943-164	Sequence 164, App
30	359	12.4	476	11	US-09-904-462-164	Sequence 164, App
31	359	12.4	476	11	US-09-907-925-164	Sequence 164, App
32	359	12.4	476	11	US-09-902-692-164	Sequence 164, App
33	359	12.4	476	11	US-09-903-520-164	Sequence 164, App
34	359	12.4	476	11	US-09-905-056-164	Sequence 164, App
35	359	12.4	476	11	US-09-909-064-164	Sequence 164, App
36	359	12.4	476	11	US-09-904-553-164	Sequence 164, App
37	359	12.4	476	11	US-09-905-381-164	Sequence 164, App
38	359	12.4	476	11	US-09-905-088-164	Sequence 164, App
39	359	12.4	476	11	US-09-907-575-164	Sequence 164, App
40	359	12.4	476	11	US-09-905-075-164	Sequence 164, App
41	359	12.4	476	11	US-09-902-759-164	Sequence 164, App
42	359	12.4	476	11	US-09-902-634-164	Sequence 164, App
43	359	12.4	476	11	US-09-902-713-164	Sequence 164, App
44	359	12.4	476	11	US-09-907-979-164	Sequence 164, App
45	359	12.4	476	11	US-09-902-615-164	Sequence 164, App

ALIGNMENTS

RESULT 1
US-09-420-785A-4
; Sequence 4, Application US/09420785A
; Patent No. US20010010923A1
; GENERAL INFORMATION:
; APPLICANT: MORTENSEN, UFFE
; APPLICANT: OLESEN, KJELD
; APPLICANT: STENNICKE, HENNING
; APPLICANT: SORESENSEN, STEEN B.
; APPLICANT: BREDDAM, KLAUS
; TITLE OF INVENTION: MODIFIED CARBOXYPEPTIDASE
; FILE REFERENCE: 089187/0109
; CURRENT APPLICATION NUMBER: US/09/420,785A
; CURRENT FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-420-785A-4

Alignment Scores: 4.76e-35 Length: 421
Pred. No.: 4.76e-35 Length: 421

Score: 517.50 Matches: 142
Percent Similarity: 43.98% Conservative: 70
Best Local Similarity: 29.46% Mismatches: 185
Query Match: 17.82% Indels: 85
DB: 9 Gaps: 17

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-420-785A-4 (1-421)

```
QY 97 CGGTACAGAAACCCGGGCGAGGCGCTCGGAGACTACCGCGGTGTCAAATCTAC 156
Db 1 LysileLysAspProLysIleLeuGlyIle -----AspProAsnValThrGlnTyr 17
QY 157 TCTGGATATCGACACTCTCCCGAG- --TCACATACCTTCTCTGGTTCTTCGAAGCC 213
Db 18 ThrGlyTyrLeuAspValGluAspGluAspLysHisPhePheThrPheGluSer 37
QY 214 AGCATATACCCAGAACTGCACCTATCATCTGTGGTTGAATGGTGGCCCTGGAAGCAT 273
Db 38 ArgAsnAspProAlaLysAspProValIleLeuTrpLeuAsnGlyGlyProGlyCysSer 57
QY 274 TCTTTGATCGTCTCTCGAAGAGTGGCGCCTTGCCATGTCATTCGACTTTTGTATGAC 333
Db 58 SerLeuThrGlyLeuPhePheGluLeuGlyProSerSerIleGlyProAspLeuLysPro 77
QY 334 TACATCAACCTCACTCGTGAACAGAGTCTCCAAATTTACTATTCCTGTCCAGCCATTG 393
Db 78 IleGlyAsnProTyrSerTrpAsnSerAsnAlaThrValIlePheLeuAspGlnProVal 97
QY 394 GGATCGCGTTTCATATAGTATACGTTGATGGGTCCATTAACCTGTAAACWGGGGTC 453
Db 98 AsnValGlyPheSerTyrSer ----- 104
QY 454 GTCGAATTCGAGCTTTCGAGGAGTTCAGGCGCGGTACCCACCACTGATGCCACTGTG 513
Db 105 -----GlySerSerGly ----- 108
QY 514 ATCGATATACATCTTCGCGAGAGCGCGCTTGGGAGATCTCTGCAAGGATTCCTTAGT 573
Db 109 ValSerAsnThrValAlaAlaGlyLysAspValTyrAsnPheLeuGluLeuPhePheAsp 128
QY 574 GGACTACTAGCTGGACTTAGGTGCAGTCTAAGACTTCAGTCTATGACGAGGAGC 633
Db 129 GlnPheProGluTyrValAsnLysGlyGln -----AspPheHisIleAlaGlyGluSer 146
QY 634 TATGAGGAGGACTATGCTCTGCAATCTTCAATCATCTTTACGAGCAGAAAGAGAGATT 693
Db 147 TyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeuSerHisLysAspArg --- 165
QY 694 GCCAAGGTAGTGTATATGGTGTTCAGCTTAATTTCACTCTCTGGGAATTAATT ----- 747
Db 166 -----AsnPheAsnLeuThrSerValLeuIleGly 175
QY 748 AACGGCATATCGACGAGGCGATCCAGGCCCTTACTACCCCTGATTCGCT ----- 798
Db 176 AsnGlyLeuThrAspProLeuThrGlnTyrAsnTyrTyrGluProMetAlaCysGlyGlu 195
QY 799 -----GTGAACAATACCTACGCTATCAAGCTGTCAACGAGACCGCTCTAC 843
Db 196 GlyGlyGluProSerValLeuProSerGluGluCysSerAlaMetGluAspSerLeu --- 214
QY 844 AACTACATGAAGTTTGGCAACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
Db 215 -----GluArgCysLeuGlyLeuIleGluSerCys 224
QY 904 AAACAGACAAACCGCACCGCATTAGCTGACTACGCCCTCTGGCGGAAGCCACCAACATG 963
Db 225 TyrAspSerGln -----SerValTrpSerCysValProAlaThrIleTyr 239
QY 964 TCCAGGGAACAATGTTAGGGGCGCATACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1023
Db 240 CysAsnAsnAlaGlnLeuAlaProTyrGlnArg ---ThrGlyArgAsnValTyrAspIle 258
QY 1024 CGGCATCCATATGATGAC -----CCGACTCCGCCCAAGTTATTACACANA 1068
```

```
Db 259 ArgLysAspCysGluGlyGlyAsnLeuCysTyrProThrLeuGlnAsp ---IleAspAsp 277
QY 1069 TTTCTGGCAAGGACTCTGTCTATGGAGCTATCGCGCTCAACACTCAACTCAACACTCC 1128
Db 278 TyrLeuAsnGlnAspTyrValLysGluAlaValGlyAlaGluValAspHisTyrGluSer 297
QY 1129 ---AATAATGACGCTACTACGCTTTCACGCAACAGGAGGAGCTTTGCTGCGCCCAACTTC 1185
Db 298 CysAsnPheAspIleAsuAlaGlnPheLeuPheAlaGlyAspTrpMetLysPro ---Tyr 316
QY 1186 ATCGAAGACCTCGAGGAGATCTTCTCTCCCTCGGCTGCTCCCTCATCTATGCGGAC 1245
Db 317 HisThrAlaValThrAspLeuLeuAsnGlnAspLeuProIleLeuValTyrAlaGlyAsp 336
QY 1246 GCCATATCATCTCAACTGCTGCGCGGTTCAGGCGGTTTCCCTCGCTGCGAACTACTCC 1305
Db 337 LysAspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrAspValLeuProTrpLys 356
QY 1306 CAAGCCGCCAGTTCGGAAGC -----GCAGGGTACAGCGCCCTGAAAGTCAACGGC 1356
Db 357 TyrAspGluGluPheAlaSerGlnLysValArgAsnTrpThrAlaSerIleThrAspGlu 376
QY 1357 GTCAGATATGGGAACTCGGAGTATGTAATTTCTCTTCACTCGGCTATGAGGCA 1416
Db 377 ValAla ---GlyGluValLysSerTyrLysHisPheThrTyrLeuArgValPheAsnGly 395
QY 1417 GGCCATGAAGTCCCATCTACTACGCGCATCCCTCGCTCAATTTTAAACGGGACTATC 1476
Db 396 GlyHisMetValProPheAspValProGluAsnAlaLeuSerMetValAsnGluTrpIle 415
QY 1477 TTCGGT 1482
Db 416 HisGly 417
```

RESULT 2

US-09-901-252-15
; Sequence 15, Application US/09901252
; Patent No. US20020026658A1
; GENERAL INFORMATION:
; APPLICANT: Chagple, Clint
; TITLE OF INVENTION: Genes Encoding Sinapoylglucose:Malate Sinapoyltransferase and
; TITLE OF INVENTION: Use
; FILE REFERENCE: N1422-004
; CURRENT APPLICATION NUMBER: US/09/901,252
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Saccharomyces
US-09-901-252-15

Alignment Scores:
Pred. No.: 4,76e-35 Length: 421
Score: 517.50 Matches: 142
Percent Similarity: 43.98% Conservative: 70
Best Local Similarity: 29.46% Mismatches: 185
Query Match: 17.82% Indels: 85
DB: 9 Gaps: 17

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-901-252-15 (1-421)

```
QY 97 CGGTACAGAAACCCGGGCGAGGCGCTCGGAGACTACCGCGGTGTCAAATCTAC 156
Db 1 LysileLysAspProLysIleLeuGlyIle -----AspProAsnValThrGlnTyr 17
QY 157 TCTGGATATCGACACTCTCCCGAG- --TCACATACCTTCTCTGGTTCTTCGAAGCC 213
Db 18 ThrGlyTyrLeuAspValGluAspGluAspLysHisPhePheThrPheGluSer 37
```

QY	214	AGACATAACCCAGAAATGCACATATCACATTTGTGTTGAATGGNGGCCCTCGAAGCGAT	273
Db	38	ArgAsnAspProAlaLysAspProValIleLeuTrpLeuAsnGlyGlyProGlyCysSer	57
QY	274	TCATTGATCGGTCCTTCGGAAGAGTGGGCCCTGCCATGTCACATTCGACTTTTCATGAC	333
Db	58	SerLeuThrGlyLeuPhePheGluLeuGlyProSerSerIleGlyProAspLeuLysPro	77
QY	334	TACATCAACCCCTCCTCTGGACACAGTCTCCCAATTTACTATTCCTGTCCACGCATG	393
Db	78	IleGlyAsnProTyrSerTrpAsnSerAsnAlaThrValIlePheLeuAspGlnProVal	97
QY	394	GGAGTCGGCTTTTCATATAGTGAACGGTTGATGGTCCATTAAACCCGTAACTGGCGTC	453
Db	98	AsnValGlyPheSerTyrSer	104
QY	454	GTCGAATAATTCGAGTTTGCAGGATTCAGGCGCGGTACCAACAACATTCATGCCACTGTG	513
Db	105	-----GlySerSerGly-----	108
QY	514	ATCGATACTACCAATCTTTGCCGACAGAGCCGCTTCGGAGATCCTGCAAGGATTCCTTAGT	573
Db	109	ValSerAsnThrValAlaAlaGlyLysAspValTyrAsnPheLeuGluLeuPhePheAsp	128
QY	574	GGACTACCTAGCTTGGACTCTAGGTCGAGCTAAGGACTTCAGTCTATGCAACGAGAGC	633
Db	129	GlnPheProGluTyrValAsnLysGlyGln-----AspPheHisIleAlaGlyGluSer	146
QY	634	TATCGAGGGCACTATGGCTTCGCAATCTTCAATCATTTTACGACGAGAATGAGAGAATT	693
Db	147	TyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeuSerHisLysAspArg--	165
QY	694	GCCAAACGCTAGTGTAAATGGTGTTCAAGCTTAATTTCAACTCTCTCGGAAATATT	747
Db	166	-----AsnPheAsnLeuThrSerValLeuIleGly	175
QY	748	AACGGCATATPCGACGAGCGGATCCAGGCCCTTACTACCGTGAATTCGCT	798
Db	176	AsnGlyLeuThrAspProLeuThrGluTyrAsnTyrTyrGluProMetAlaCysGlyGlu	195
QY	799	-----GTGACAACTACCTACGGTATCAAGGTGTCAACGAGACCGCTCTAC	843
Db	196	GlyGlyGluProSerValLeuProSerGluGluCysSerAlaMetGluAspSerLeu--	214
QY	844	AACTACATGAAGTTTGCCAAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGC	903
Db	215	-----GluArgCysLeuGlyLeuIleGluSerCys	224
QY	904	AAACAGACAAACCGCACCGCATAGCTAGCTAGCCCTCTCCGCCGAGGCACCAACATG	963
Db	225	TyrAspSerGln-----SerValTyrSerCysValProAlaThrIleTyr	239
QY	964	TCGAGGACAAATGTTGGGGCCCATACTACGCCCTTCTGGTCGTGTGTGTGTGTGTGTGT	1023
Db	240	CysAsnAsnAlaGlnLeuAlaProTyrGlnArg---ThrGlyArgAsnValTyrAspIle	258
QY	1024	CGGCATCCCATATGATGAC-----CCGACTCCGCCCAAGTTATTACACACAAA	1068
Db	259	ArgLysAspCysGluGlyGlyAsnLeuCysTyrProThrLeuGlnAsp---IleAspAsp	277
QY	1069	TTTCTGCGCAAGGACTCTGTCATGGAGCTATCGGCGTCAACTCAACTACACCCAGTCC	1128
Db	278	TyrLeuAsnGlnAspTyrValLysGluAlaValGlyAlaGluValAspHisTyrGluSer	297
QY	1129	---AAATAATCAAGTCTACTACGCTTTTCCACGAACAGCGCACTTTGTCTGGCCCCAATTC	1185
Db	298	CysAsnPheAspIleAsnArgAsnPheLeuPheAlaGlyAspTrpMetLysPro---Tyr	316
QY	1186	ATCGAAGACCTCGAGGAGATCCTTGCTCTCCCGCTGGTGTCTCCCTCATCTATGCGCAC	1245
Db	317	HisThrAlaValThrAspLeuLeuAsnGlnAspLeuProIleLeuValTyrAlaGlyAsp	336

```

QY      1246 GCCGATTACATCTGCACAAGTGGTCGGCGGTCCGGCCTGAGGCCGTTTCCTCGTGCGAATACTACTCC 1303
Db      |||||:::|||||:::||| ||| ::||| ::|
QY      337 LysAspPheIleCysAsnTrpLeuGlyAsnLysAlairThrAspValLeuProTrpIlys 356
QY      1306 CAAGCGCCCACTTCCGAAGC-----GCAGGGTACAGCCCCCTGAAAGTCAACGGC 1356
Db      ::||| ||| ::||| ::||| ::||| ::|||
QY      357 TyrAspGluGluPheAlaserGlnLysValArgAsnTrpThrAlaSerIleThrAspGlu 376
QY      1357 CTCGAGTAGTGGGAACATCGCGAGTATGTAATTCCTCCCTCACICGCGTCAANGAGCCA 1416
Db      ||| ||||| ::| ||| ::|||::||| |||||::|
QY      377 ValAla---GlyGluValLysSerTyrLysHisPheThrTyrLeuArgValPheAsnGly 395
QY      1417 GGCATGAAGTCCCATACTACACGCCCATCGCCCTCGCAATGTGTTAACCGGACTATC 1476
Db      ||||| |||||::| ||| ::||| ::| ||| |||
QY      396 GlyHisMetValProPheAspValProGluAsnAlaLeuSerMetValAsnGluTrpIle 415
QY      1477 TTCGGT 1482
Db      |||
Db      416 HisGly 417

RESULT 3
US-10-084-018-3
; Sequence 3, Application US/I0084018
; Publication No. US20020160499A1
; GENERAL INFORMATION:
; APPLICANT: Handman, Olga
;              Hawkins, Phillip R.
;              Hillman, Jennifer L.
;              Lal, Preeti
;              Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN SERINE
;                   CARBOXYPEPTIDASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/I0/084,018
; FILING DATE: 25-Feb-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,689A
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/828,488
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0241 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MMLR3DT01
; CLONE: 566993
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-084-018-3

Alignment Scores:
```

Pred. No.:	8.72e-22	Length:	476
Score:	361.00	Matches:	130
Percent Similarity:	39.51%	Conservative:	62
Best Local Similarity:	26.75%	Mismatches:	178
Query Match:	12.43%	Indels:	116
DB:	14	Gaps:	21
US-09-712-338-I_COPY_55_1662 (1-1608) x US-10-084-018-3 (1-476)			
QY	139	CGGGT-----GTCAAAATCTACTCTGGATAT-----GTCAACACCTCTCCGAGTCC	186
DB	67	ProGlyLeuAsnMetIysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer	86
QY	187	CATACCTTCTTCGGTCTTCGAAGCAGACATACCCAGAACTGCACCTATCACTG	246
DB	87	AsnLeuPhePheThrPheProAlaGlnLeuGlnProGluAspAlaProValValLeu	106
QY	247	TGGTTGAATGGTGGCCCTGGAACGATTTCTTCATCGGCTCTTCCAAAGAGTGGCCCT	306
DB	107	TrpLeuGlnGlyGlyProGlyGlySerSerMet***GlyLeuPheValGluHisGlyPro	126
QY	307	TGCCATGTCAATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGACAGAGTCTCC	366
DB	127	TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProIhrThrThr***	146
QY	367	AATTTACTATTCCTGTCCCGACCATTTGGGAGTCGGCTTTTCAATAGTAGTCGGTTGAT	426
DB	147	SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis	166
QY	427	GGGTCCATTAACTGTAACTGGGTCGTCGAAAATTCGAGCTTTCAGGAGTTCAGGC	486
DB	167	Gly-----	167
QY	487	CGGTACCCAAACCAATTGATGACTCTGATGACTACTACCAATCTGCCGAGAGCGCT	546
DB	168	-----TyrAlaValAsnGluAspValAlaArgAspLeuTyr	180
QY	547	TGGGAGATCCCTGCAAGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGAGCACTC	606
DB	181	SerAlaLeuLeuGlnPhePhe-----GlnIlePheProGluTyrLysAsn	195
QY	607	AAGACATTCAGTCTATGCGAGCAGACTATGGAGGCACCTATGCTCGCTCATTTCAAT	666
DB	196	AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis	215
QY	667	CATTTTACGACAGAAATGAGAAATTCGCAACGGTAGTGTAAATGGTGTACGTTAA	726
DB	216	LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn	229
QY	727	TTCACTCTCWGGAAATTTAACCAGCATCAICGACGAGCGATCAGGCCCTTACTAC	786
DB	230	LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr	249
QY	787	CCTGAATTCGCTGGAACAATCACTACGGTATCAAGGCTGTCAAGCAGCGCTACAAC	846
DB	250	AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys	265
QY	847	TACATGAAGTTTGCCAAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACTGCAA	906
DB	266	Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg	279
QY	907	CAGACAAC-----CGCACCGCATTAGCTGACTACGCCCTCTCGCGCGAGCCACC	957
DB	280	LysGlnAsnIhrPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr	299
QY	958	AAC-----ATGTGACGGACAATTTGTAGGGG-----CCATACTACGCTTGTGGT	1005
DB	300	SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu---	318
QY	1006	CGTGGTGTATGATATTCGGCATCCATATGATGACCCGACTCGCCCAAGTATTATCAAC	1065
DB	319	-----ArgCysThrGluProGluAspGlnLeuTyrTyrVal	330

QY	487	CGGTACCCAAACATTGATGCCACTCTGATCGATACCTACCAATCTTGCGCGACAGCGCCGCT	548
Db	168	-----TyrAlaValAsnGluAspValAlaAlaArgAspLeuTyr	180
QY	547	TGGGAGATCTCCAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGTCGAGTCT	606
Db	181	SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn	195
QY	607	AAGGCTTCAGTCTATGGACGGAGAGACTATGGAGGGCACTATGCTGCTCGCACTTCTCAAT	666
Db	196	AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis	215
QY	667	CATTTTACGACAGAATCAGAGATTCGCCAACGGTACTGTTAATGTGTGTTCACTTAAT	726
Db	216	LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn	229
QY	727	TTCACACTCTCGGGAATTAATTAACGGCATCATCGACGAGCGATCCAGGCCCTTACTACT	786
Db	230	LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr	248
QY	787	CGTGAATTCGCTGTGAACAATACCTACGGGTATCAAGGCTGTCAACGACGACCGCTACAAC	846
Db	250	AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys	265
QY	847	TACATGAAGTTTGCAACCAATGCAAAAGTTGCCAGGATTTGATTCCACCTGCAGAA	906
Db	266	Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg	279
QY	907	CAGACAAAC-----CGCACCGCAATTACTGACTACGCGCTCTGCGGGAAGCAACC	957
Db	280	LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr	299
QY	958	AAC-----ATGTCAGGACAATGTCAGGG-----CCATACTACGCGCTTGCTGGT	1005
Db	300	SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu---	318
QY	1006	CGTGTGTGTATGATATTCGCACTCCATATGATACCGCGACTCGGCAAGTTATTACAAC	1065
Db	319	-----ArgCysThrGluProGluAspGlnLeuTyrTyrVal	330
QY	1066	AAATTTCTGGCAAGGACTCTGTATGGAGCTATGCGGCTCAACATCACTACACCCAG	1125
Db	331	LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe	349
QY	1126	TCCAATAATGACGCTAC-----IACGCTTCACCAACAGCGCACTTGCTGCGCC	1179
Db	350	AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro	369
QY	1180	AACTTCATCGAAGACCTCGAGGAGATCCTTGCTCCCGCTGGTGTCTCCCTCATCTAT	1239
Db	370	TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn	384
QY	1240	GGCACCGCATATCATC-----TGCAAC	1263
Db	385	GlyGlnLeuAspIleIleValAlaAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp	404
QY	1264	TGTTCTCGCGGTGAGGCCGTTCCCTCGCTGCGAATCTCTCCCAAGCCGCCAG-----	1317
Db	405	TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal	418
QY	1318	-----TTCGGAAC-----GCAGGTACACGCCCTCGAAAGTCAACGCG	1356
Db	419	TrpLysIlePheLysSerAspSerGluValAlaGlyTyr-----	431
QY	1357	GTCCAGTATGGGAAACTCGCGATATGGTAATTTCTCTCTCTACTCGCTCATGAGCA	1416
Db	432	-----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly	446
QY	1417	GGCATGAAGTCCCATACTACACGCCATCGCTCTCTGCAATTTGTTAACCGGCAATATC	1476
Db	447	GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle	466
QY	1477	TTC-----GGTTGGGAT	1488


```
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
QY 1006 CGTGTGTGTGAIAITCGGCATCCATATGATGACCGGACTCGCGCAAGTTATACAA 1065
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
QY 1066 AATTTCTGCAAGGACTGTGTCATGAGCGCTATCGCGGTCAACAATCAACATACAC 1125
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 TCCAAATATACGCTATC-----TAGCTTTCCAGCAACAGCGGACTTTGTCGCC 1179
Db 350 AsuAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLys 369
QY 1180 AACTTTCATGCAAGACGCGAGAGATCCITGCTCCCGGTGCGTCCCTCATCTAT 1239
Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384
QY 1240 GCGACGCGCATATATC-----TGCAAC 1263
Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY 1264 TGCTTGGCGGTGAGCGCGTTCCCTCGCTGCGAATCTCCCAAGCGCGCCAG----- 1317
Db 405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY 1318 -----TTCCGAAGC-----GCAGGTACACGCCCTCGAAGTCAACGGC 1356
Db 419 TrpLysIlePheLysSerAspSerGluValAlaIleTyr----- 431
QY 1357 GTGAGTATGGGAAACTCGGAGTATGGTAATTCCTTCCTACGCGGCTATGAGGCA 1416
Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY 1417 GGCCATGAATCCATACACTACGACCGACCGCTCCCTCGAATGTTTACCGGATATC 1476
Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY 1477 TTC-----GGTGGGAT 1488
Db 467 IyrGlyLysGlyTrpAsp 472
```

RESULT 6

```
US-09-905-291A-164
; Sequence 164, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
```

```
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRI
; ORGANISM: Homo sapiens
; US-09-905-291A-164

Alignment Scores:
Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: 10 Gaps: 21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-905-291A-164 (1-476)

QY 139 CCGGGT-----GTCAATCCCTACTCTGGATAT-----GTGACACCTCTCCCGAGTCC 186
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY 187 CATACCTCTCTGTTCTTCGAAGCCACACATACCCAGAACTCCACCTATCATCATG 246
Db 87 AsnLeuPhePheTrpPheProAlaGlnIleGlnProGluAspAlaProValValLeu 106
QY 247 TGGTTGAATGTGGCCCTGGAAGCATCTTTGATCGTCTCTTCAAGAGTGGCCCT 306
Db 107 TrpLeuGlnGlyProGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
QY 307 TCCCATGTCAATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGAGGCTCC 366
Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProThrThrThrLeu 146
QY 367 AATTACTAATCTCTCCAGCCATTGGGAGTCGGCTTTTTCATATAGTATACGGTTGAT 426
```

Db	147	SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis	166
Qy	427	GGGTCCATTAAACCTTGTAACCTGGTCTCGAATATCGACCTTGCAGGAGTTCAGGGC	486
Db	167	Gly-----	167
Qy	487	CGGTACCCCAACCATTTGATGCCACTCTGATCGATACCTACCAATCTTGGCGAGAGCGCGCT	546
Db	168	-----TyrAlaValAsnGluAspValAlaArgAspLeuTyr	180
Qy	547	TGGAGATCCTGCAAGATTCCTTAGTGAGTACCTAGCTGGACTCTAGGAGTGCAGCTCT	606
Db	181	SerAlaLeuIleGlnPhe-----GlnIlePheProGluTyrLysAsn	195
Qy	607	AAGACTTCAGTCTGAGGAGGAGCTATGGAGGCGACTATGGCTCGATCTCTCAAT	666
Db	196	AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis	215
Qy	667	CAUTTTTACGACAGATGAGAGAAATGCCACGGTAGTGTGTTAATGGTGTTCAGCTTAAT	726
Db	216	LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn	229
Qy	727	TTCACATCTCTGGGAATATTAAAGGCATCATCGAGGAGCGATCCAGCGCCCTTACAC	786
Db	230	LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr	249
Qy	787	CTGGAATCTCTGGAACATACCTACGGTATCAAGCTGTCAAGAGAGACCGTCTACAAC	846
Db	250	AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys	265
Qy	847	TACATGAAGTTTGCCAAACCAATGCCAAATGGTGGCAGGATTTGATTTCCACCTGCCAA	906
Db	266	Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg	279
Qy	907	CAGACAAC-----CGACGGCATTAAGTCAGTACGCCCTCTCGCGCGAGCCACC	957
Db	280	LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr	299
Qy	958	AAC-----ATGTCCAGGACAATGTGAGGG-----CCATACCTACGCTTTGCTGT	1005
Db	300	SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu	318
Qy	1006	CGTGTGTGTATGATATTCGGCATCCATATGATGACCCGACTCCGCCAAGTTATTACAC	1065
Db	319	-----ArgCysThrGluProGluAspGlnLeuTyrTyrVal	330
Qy	1066	AAATTTGCGCAAGGACTCTGTATGAGCGCTATCGCGTCAACATCACTACACCCAG	1125
Db	331	LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly-----AsnGlnThrPhe	349
Qy	1126	TCCATATAGCTCTAC-----TAGCTTTCAGCAAAACAGCGCGACTTTGCTGCGCC	1179
Db	350	AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro	369
Qy	1180	AACTTCATCGAAGACCTCGAGGAGATCTTCTCTCCCGTGGCGTCTCCCTCATCTAT	1239
Db	370	TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn	384
Qy	1240	GCGACGCGGATATAC-----TGCAAC	1263
Db	385	GlyGlnLeuAspIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp	404
Qy	1264	TGTTTCGGGGTTCAGCGCGTTTCCCTCGTCCGAACTACCTCCCAAGCGCGCCAG-	1317
Db	405	TrpLysGlySerGln-----GluTyrLysLysAlaGluLysVal	418
Qy	1318	-----TTCGGAAGC-----GCAGGGTACAGCCCTCGAAGAGTCAACGGC	1356
Db	419	TrpLysIlePheLysSerAspSerGluValAlaGlyTyr-----	431
Qy	1357	GTGAGTAGTGGGAACACTCGCAGATGTGTAATTTCTCTTCACTCGGCTCATGAGCA	1416

; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 164
 ; LENGTH: 476
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 ; US-09-902-853-164

Alignment Scores:

Pred. No.: 1,29e-21 Length: 476
 Score: 359.00 Matches: 130
 Percent Similarity: 39.51% Conservatives: 62
 Best Local Similarity: 26.75% Mismatches: 178
 Query Match: 12.36% Indels: 116
 DB: 10 Gaps: 21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-902-853-164 (1-476)

```

QY 139 CCGGGT-----GTCAAATCCTACTCTGGATAT-----GTGACACCTCTCCGAGTCC 186
    ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 187 CATACCTCTCTCGTTCGAGCCAGCAGACATACCCAGAACTGCACATATACATG 246
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 87 AsnLeuPheThrPheProAlaGlnLeuProGluAspAlaProValValLeu 106
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 247 TGGTCAATGGTGGCCCTGGAGCGATCTTTCATCGGTCTCTCGAAGAGTTGGCCCT 306
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 107 TrpLeuGlnGlyGlyProGlyGlySerMetPheGlyLeuPheValGluHisGlyPro 126
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 307 TGGCATGTCAATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGACGAGGTCTCC 356
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 127 TyrValValThrSerAsnMetThrLeuArgAspPheProThrThrThrLeu 146
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 367 AATTTACTATCTCTCCCGCCATTTGGAGTCGGCTTTTCATATAGTCATACGGTTGAT 426
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 147 SerMetLeuThrIleAspAsnProValGlyThrGlyPheSerPheThrAspAspThrHis 166
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 427 GGTGTCATTAACCTCTAACTGGGGTCTGTCGAAATTCGAGCTTTCGAGAGTICAGGCG 486
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 167 Gly----- 167
QY 487 CGGTACCCCAACCATTCATGCCACTCTGATCGATACGATACCAATCTTCGCGAGGCGCT 546
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 168 -----TyrAlaValAsnGluAspValAlaArgAspLeuTyr 180
QY 547 TGGGAGATCCTGCAAGGATTCCTTACTGGAGTACCTAGCTTGGACTTACGGTGCAGTCT 606
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
QY 607 AAGGACTTCAGTCTATGGACGAGAGCTATGAGGCGCACTATGCTTCCTCATTCAT 666
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 196 AsnAspPheTyrValThrGlyGlySerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY 667 CATTTTACGACAGAAAGAGAAATTCGCAACGGTGTGTTAAGTGTTCAGCTTAAT 726
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn 229
QY 727 TTCAACTCTCTGGGAATTATTAAACGSCATCATCGAGGCGCATCCAGGCCCTTACTAC 786
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249
QY 787 CCGTAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACCGCTCTACAC 846
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265
QY 847 TACATGAAGTTTGGCAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGCAA 906
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  
```

```

Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY 907 CAGACAAAC-----CGCACCGCATTAGCTAGCTACGCGCTCTGCGCGCAAGCCACC 957
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299
QY 958 AAC-----ATGTCAGGACAAATGTTGAGGG-----CCATACIACGCCCTTGTCTGG 1005
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
QY 1006 CGTGGTGTGTATGATATTCGGCAITCCATATGATACCGGACITCCGCCAAGTTATACAC 1065
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
QY 1066 AATTTCTGGCAAGAGACTCTGTCTATGACGCTATCGGCGCTCAACATCACTACACAG 1125
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 ICCAATAATGACGTCTAC-----TAGCTTTCAGCAACACAGGCGACTTTGTCTGGCC 1179
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY 1180 AACTTCATCGAAGACCTCGAGGAGATCCTTGTCTCTCCCGTGCCTGCTCCTCAICTAT 1239
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384
QY 1240 GCGCACGCCGATATATC-----TGCACAA 1263
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY 1264 TGGTTCGGCGTACGCGCTTTCCTCGCTGCGAAGTACTCCCAAGCGCGCCAG----- 1317
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysVal 418
QY 1318 -----TTCGCAAGC-----GCAGGTACACGCGCCCTGAAGTCAACGCGC 1356
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY 1357 GTCAGTATGGGAAACTCGCGAGTATGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1416
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY 1417 GGCATCAAGTCCCATCTACTACCGCCATCGCTCCCTCGCAATGTGTTTACCGGACTATC 1476
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY 1477 TTC-----GGTTGGGAT 1488
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 467 TyrGlyLysGlyTrpAsp 472

```

RESULT 8

US-09-907-824-164
 ; Sequence 164, Application US/09907824
 ; Publication No. US20020197671A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: Kljavin, Ivar J.
 ; APPLICANT: Mather, Jeannie P.

```

; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,824
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-907-824-164

Alignment Scores:
Pred. No.: 1,29e-21
Score: 359.00
Percent Similarity: 39.51%
Best Local Similarity: 26.75%
Query Match: 12.36%
DB: 10

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-907-824-164 (1-476)
Qy 139 CCGGGT-----GTCAATCTCTGGATAT-----GTCGACCTCTCCGAGTCC 186
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
Qy 187 CATACCTCTCTGGTCTTCGAGCCAGACATACCGAGAACTGCACCTATCACATTG 246
Db 87 AsnLeuPhePheTrpPhePheProAlaGlnIleGlnProGluAspAlaProValValLeu 106

```

```

QY 247 TGGTTGAATGGTGGCCCTCGAAGCGGATCTTTTCATCGGTCTCTTCAAGAGTTGGGCCCT 306
Db 107 TrpLeuGlnGlyProGlySerMetPheGlyLeuValGluHisGlyPro 126
QY 307 TGGCATGTCAATTCGACATTTTGATGACTACATCAACCTCTCTCGGTGGAACGAGTCTCC 366
Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProIleThrThrLeu 146
QY 367 AATTTTACTATTCCTGTCGCCCAATTTGGGAGTCGGCTTTTTCATATAGTATAGATCGGTGAT 426
Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis 166
QY 427 GGGTCCAITAACCCCTGTAACIGGGGTCTGCGAAATTCGAGCTTTCGAGGAGTTCAGGGC 486
Db 167 Gly----- 167
QY 487 CGGTACCCACCAATTGATGCGACTCTGATCGTACTACCAATCTTCCGCGAGAGCGGCT 546
Db 168 -----TyrAlaValAsnGluAspValAlaArgAspLeuTyr 180
QY 547 TGGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTGCAGTCT 606
Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
QY 607 AAGGACTTCACTCTATGGACGAGAGCCTATGGGGGACACTATGGTCTCGATCTTCAAT 666
Db 196 AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY 667 CATTTTACGAGCAGAGATGAGAAATGCCAAGGTAGTGTATATGGTGTTCAGCTTAAT 726
Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn 229
QY 727 TTCAACTCTCTGGGAATTTAAGCGCATCATGACGAGCGGATCCAGGCCCTTACTAC 786
Db 230 LeuAsnGlyIleAlaIleGlyAspGlyIyrSerAspProGluSerIleIleGlyIyr 249
QY 787 CCTGAATTCGTGTGAACATACCTACGGTATCAAGGCTGTCACAGACACGCTACAAC 846
Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265
QY 847 TACATGAAGTTTGCCAAATGCCAAATGGTTGCCAGGATTTGATTTCCACCTGCAAA 906
Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY 907 CAGACAAAC-----CCGACCGCATTAAGTACGCGCTCTCGCGCGAGCCACC 957
Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299
QY 958 AAC-----ATGTCGAGGACAATGTTGAGGG-----CCATACCTAGCCCTTGTCTGT 1005
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
QY 1006 CGTGGTGTGTATGATATTCGGCATCCATATGATGACCGGACTCCGCCCAAGTATTACAAAC 1065
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
QY 1066 AATTTCTGCGAAGACTCTGTATGAGCGCTATCGCGCTCAACATCACTACACCCAG 1125
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 TCCAATATGAGTCTAC-----TACGCTTTCAGCAACACAGCGGACTTTGTCTGGGCC 1179
Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY 1180 AACTTCATCGAAGACTCGAGGAGATCCTTGTCTCTCCCGTGGCTGTCTCCCTCATCAT 1239
Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384
QY 1240 GCGACGCGGATATCATC-----TGCAC-----TGCAC 1263
Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY 1264 TGGTTCGCGGTGACGCGCTTTCCTCGCTGCGAACTACTCTCCCAAGCGGCCAG----- 1317

```

```

Db      405  TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY      1318  -----TTCGGAAG-----GCAGGGTACAGCCCTCAAGTCAACGGC 1356
Db      419  TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY      1357  GTCAGATATGGGAACCTCGCGAGTATGGTAATTTCTCTCACTCGCGTCTATGAGGCA 1416
Db      432  -----IleArgGluAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY      1417  GCCCATGAAGTCCCTACTACCGCCATCCGCTCCCTGCAATGCTTTAACCGGACTATC 1476
Db      447  GlyHisIleLeuPzTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY      1477  TTC-----GGTGGGAT 1488
Db      467  TyrGlyLysGlyTyrPasp 472

```

RESULT 9

```

US-09-907-841-164
; Sequence 164, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214

```

```

; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-841-164

```

Alignment Scores:

```

Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: 10 Gaps: 21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-907-841-164 (1-476)

QY      139  CCGGGT-----GTCATATCTCTCTGGATAT-----GTGACACCTCTCCCGAGTCC 186
Db      67  ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY      187  CATACCTTCTCTGTTCTTCGAGAGCCAGACAIACCAGAACTGCACCTATCACATTG 246
Db      87  AsnLeuPhePheTyrPhePheProAlaGlnIleGlnProGluAspAlaProValValLeu 106
QY      247  TGGTTGAATGGTGGCCCTCGAAGGATTCITTTGATCGGTCTCTTCGAAAGAGTGGGCC 306
Db      107  TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
QY      307  TGCCATGTCAATTCGACTTTTGTATGACACATCAACCTCTACTCGTGGAGAGGTCCTCC 366
Db      127  TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProTyrThrThrLeu 146
QY      367  AATTTACTATTCTCCGTCACGCCCATTCGGAGTCGGCTTTTCATATAGTATACGGTTGAT 426
Db      147  SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis 166
QY      427  GGGTCCATTAAACCCCTGTAACCTGGGGTCTGCGAAATTCGAGCTTTGAGGAGTTCAGGG 486
Db      167  Gly----- 167
QY      487  CGGTACCCACCATTCATGATGCCACTCTGATCATCTACTACCAATCTGCGAGAGCGCGCT 546
Db      168  -----TyrAlaValAsnGluAspValAlaAlaArgAspLeuTyr 180
QY      547  TGGGAGATCCTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTCTAGGGTCACTCT 606
Db      181  SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
QY      607  AAGGACTTCAGTCTATGGACGAGAGCTATGAGGSCACTATGCTCCTGCATTTCTCAAT 666
Db      196  AsnAspPheTyrValThrGlyGlySerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY      667  CAITTTTACGACAGATGACAGAAATGCCAAGGTAGTCTTATGTTGCTGCTACGTTAAT 726
Db      216  LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn 229
QY      727  TTCAACTCTCTGGGAATTAATAACGGCATCATCGACGAGCGGATCCAGGCCCTTACTAC 786
Db      230  LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyTyr 249
QY      787  CQTGAATTCGCTCTGAACAATACCTACGGTATATCAAGCTGTCAACGAGACCGGTCTACA 846
Db      250  AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265
QY      847  TACATGAAGTTTCCCAACCAATGCCAATGGTTGCCAGGATTTGATTTCCACTCCAA 906
Db      266  Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY      907  CACACAAAC-----CGCACCGCATTAGTGTGACIACGCCCTCTGCGCGGAGCAAC 957

```

Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299
QY 958 AAC-----ATGTGAGGACAAATGTTAGGG-----CCATCTACCCCTTGTGTT 1005
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
QY 1006 CGTGGTGTCTATCATATCGGCATCCATATGATGACCCGACTCCGCCAAGTTATTATCAAC 1065
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
QY 1066 AAATTTCTGCAAGAACACTGTGTATGAGCAGTCATCGCGTCAACATCAACACACCCAG 1125
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 TCCAAATAATCAGCTCTAC-----TAGCGTTTCCAGCAACAGCGGACATTGTGGCC 1179
Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY 1180 AACTTTCATCAAGACCTCCGAGGAGATCTTCTCTCCCGTGGTGTCTCCCTCACTAT 1239
Db 370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384
QY 1240 GCGCAGCGCATATCATC-----TGCAC 1263
Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY 1264 TGGTTCGGCGGTGAGCGCGTTTCCCTCGCTCGCAACTACTCCCAAGCGCCCGAG----- 1317
Db 405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY 1318 -----TTCGGAAC-----GCAGGGTACACGCCCTCGAAAGTCAACGGC 1356
Db 419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY 1357 GTCGAGTATGGGAACACTCCGAGATATGGTAATTTCCCTTACTCCCGTCTAAGAGCA 1416
Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlygly 446
QY 1417 GGCATGAGTCCCATACTACCGCCATCCCTCGCTGCAATGTGTTTAAACGGACATC 1476
Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY 1477 TTC-----GGTTGGGAT 1488
Db 467 TyrGlyLysGlyTrpAsp 472

RESULT 10
US-09-904-011-164
; Sequence 164, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jeanie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904_011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-904-011-164
Alignment Scores:
Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: Gaps: 21
US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-904-011-164 (1-476)
QY 139 CCGGT-----GTCAATCTCTCTGTGATAT-----GTCCGACCTCTCCCGAGTCC 186
Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY 187 CATACCTTCTCTGTGTTCTCGAAGCCAGACATAACCCAGAACCTGCACCTATCATATG 246
Db 87 AsnLeuPhePheTrpPheProAlaGlnIleGlnProGluAspAlaProValValLeu 106
QY 247 TGGTTGAATGGTGGCCCTGGAAGCGATTTTGTATCGGTCTCTTCGAGAGTGGCCCT 306
Db 107 TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126

; PRIOR FILING DATE: 1999-11-30
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 164
 ; LENGTH: 476
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-09-906-742-164

Alignment Scores:

Pred. No.: 1,29e-21 Length: 476
 Score: 359.00 Matches: 130
 Percent Similarity: 39.51% Conservative: 62
 Best Local Similarity: 26.75% Mismatches: 178
 Query Match: 12.36% Indels: 116
 Gaps: 21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-906-742-164 (1-476)

QY 139 CCGGCT-----GTCAAACTCTACTCTGATAT-----GTGACACACTCTCCCGAGTCC 186
 Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
 QY 187 CAPACCTCTCTGCTTTCGAGCCAGCAGACATACCCAGAACTGCACCTATCACATTS 246
 Db 87 AsnLeuPheThrPheProAlaGlnProGluAspAlaProValLeu 106
 QY 247 TGGTGAATGGTGGCTGAGCGATTTTCATGCTCTTCGAGAGATTGGGCGCT 306
 Db 107 TriPleuGlnGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
 QY 307 TGCCATGTCAATTCGACTTTTGATGACTACATCAACCTCCTCGTGGAGAGGTCTCC 366
 Db 127 TyrValValThrSerAsnMetThrLeuArgAspPheProTyrThrThrLeu 146
 QY 367 AATTACTATCTGCCCCAGCATGGGAGTGGCTTTTCATATAGTATGATCGATTGAT 426
 Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis 166
 QY 427 GGTGCCATTACCTGTAACCTGGGTGCTCGAAATTCGAGCTTTGCAGGAGTTCAGGGC 486
 Db 167 Gly----- 167
 QY 487 CGGTACCCACCAATTGATGCCACTTCGATGCTACTACCAATCTGCGCAGAGCGGCT 546
 Db 168 -----TyrAlaValAsnGluAspValAlaArgAspLeuTyr 180
 QY 547 TGGGAGATCTGGAAGATTCCTTAGTGGACTAGCTAGCTAGCTAGCTAGCTAGCT 606
 Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
 QY 607 AAGGACTTCATCTATGACGAGGAGCTATGGAGGCTATGGTCTCTGCTATCTCAAT 666
 Db 196 AsnAspPheTyrValThrGlySerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
 QY 667 CATTTTTCAGCAGAGATGAGAGAAATGCCAACGGTAGTGTAAATGGGTTCACCTTAAT 726
 Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn 229
 QY 727 TTCACACTCTGGGAATTAATTAACGGCATCATCGACGAGGAGATCCAGGCCCTTACAC 786
 Db 230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249

QY 787 CCTGAATTCGCTGTGAACAATACTACGGTATCAAGGCTGTCAACAGACGACCTCTACAAC 846
 Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGlyLysGlnLysLys 265
 QY 847 TACATGAAGTTTCCCAACCAATGCAATGGTTGCCAGGATTTGATTTCCACCTCAAA 906
 Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
 QY 907 CAGACAAAC-----CGCACCGCATTAGCTAGCTACGCCCTCTGCGCCGAGCAACC 957
 Db 280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyLeuThr 299
 QY 958 AAC-----ATGTGACAGGACAATGTTGAGGG-----CCATACACTACGCTTTGCTGGT 1005
 Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
 QY 1006 CGTGGTGTGTATGATATTCGGCATCATATGATGACCCGACTCGGCCAACTTTATACAAC 1065
 Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
 QY 1066 AAATTTCTGGCAAGGACTCTGTCAAGCGCTATCGCGCTCAACATCAACTACACCCAG 1125
 Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
 QY 1126 TCCAATAAATGACGCTTAC-----TAGCTTTTCCAGCAACAGCGGACTTTGCTGGCC 1179
 Db 350 AsnAspGlyThrIleValGlnLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
 QY 1180 AACTTCATCAAGACCTCGAGGAGATCCTTGCCTCCCGCTGCTGCTCCCTCATCTAT 1239
 Db 370 TrpLeuThrGluLeuMetAsnAsn-----TyrLysValLeuIleTyrAsn 384
 QY 1240 GCGACGCGCATATACATC-----TGCAAC 1263
 Db 385 GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
 QY 1264 TGGTTCGGGCTAGCGCGTTTCCCTCGTCCGAACTACTCCCAAGCGGCCAG----- 1317
 Db 405 TriPysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
 QY 1318 -----TWCCGAAGC-----GCAGGTACAGCGCCCTGAAAGTCAACGGC 1356
 Db 419 TriPysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
 QY 1357 GTCGAGTATGGGAACTCGGAGTATGTAATTTCTCTCTACTCGCGTCTATCAGGCA 1416
 Db 432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
 QY 1417 GGCCATGAAGTCCCATACTACCGCCATCCCGCCCTCCCTGCAATTTGTTTAAACCGACTATC 1476
 Db 447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
 QY 1477 TTC-----GGTGGGAT 1488
 Db 467 TyrGlyLysGlyTyrAsp 472

RESULT 12

US-09-906-838-164
 ; Sequence 164, Application US/09906838
 ; Publication No. US20030027143A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.

```

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tomas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,838
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-906-838-164

```

```

Alignment Scores:
Pred. No.: 1,29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.36% Indels: 116
DB: 11 Gaps: 21

```

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-906-838-164 (1-476)

QY 139 CCGGGT-----GTCAAAATCTACTCTGGATAT-----GTGACACCTCTCCGAGTCC 186

```

Db 67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAlaSerLysThrTyrAsnSer 86
QY 187 CATACCTCTCTTGGTCTTTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTG 246
Db 87 AsnLeuPhePheIrrPhePheProAlaGlnIleGlnProGluAspAlaProValLeu 106
QY 247 TGGTGAATGGTGGCCCTGGAGCGATCTTTGTATCGGTCTCTTCGAGAGITGGCCCT 306
Db 107 TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
QY 307 TGGCATCTCAATTCGACTTTTGATGACTACATCAACCCCTCACTCTCGAGAGAGGTCTCC 366
Db 127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProIrrThrThrLeu 146
QY 367 AATTACTATTCCTGTCCTCCAGCCATTCGGAGTCGGCTTTTCATATAGTATACGGTTGAT 426
Db 147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspAspThrHis 166
QY 427 GGGTCCATTAAACCTGTAACTGGGTCTGTCGAAATTCGAGCTTTCGAGAGTTCAGGGC 486
Db 167 Gly----- 167
QY 487 CGGTACCCAAACCATTTGAIGCCACTCTGATCGATACCAATCTTGCCGAGAGGCCGCT 546
Db 168 -----TyrAlaValAlaSerLysThrValAlaArgAspLeuTyr 180
QY 547 IGGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTTGGAGTTCAGGTCT 606
Db 181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn 195
QY 607 AAGGACTTCAGTCTATGGAGCGAGAGTATGGAGGCGACATPAGTGTCTGCATTCTTCAAT 666
Db 196 AsnAspPheTyrValThrGlyGlySerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY 667 CATTTTTCAGCAGAGATGAGAGAATGCCAACGGTAGTGTATGTTGTTGCTGACGTTAAT 726
Db 216 LeuIleHisSerLeuAsn-----ProValArgGluValLysLeuAsn 229
QY 727 TTCAACTCTCTGGGAATATTATACGGCATCATCGACGAGCGATCCAGGCCCTTACTAC 786
Db 230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249
QY 787 CCGTAATTCGCTGTGAACAATACCTACGGTATCAAGCGTGTCAACGACCGCTCTACAAC 846
Db 250 AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys 265
QY 847 TACATGAAGTTTCCCAACCAATGCCAAATGGTTGCCAGSATTTGATTTCACCTGCACA 906
Db 266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY 907 CAGACAAAC-----CGCACCGCANTAGTCTACGCGCTCTGCGCGCAAGCCAGC 957
Db 280 LysGlnAsnTrpPheGluAlaPheGluLeuAspLysLeuLeuAspGlyAspLeuThr 299
QY 958 AAC-----ATGTCGAGGCGACAATGTTGAGGGG-----CCATACTACCGCTTGTGT 1005
Db 300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu--- 318
QY 1006 CGTGGTGTGTATGATATTCGGCAICCATATGATGACCGCACICCCGCAAGTTATTACAAC 1065
Db 319 -----ArgCysThrGluProGluAspGlnLeuTyrVal 330
QY 1066 AATTTCTGCAAGAGACTCTGTCTATGGACGCTATCGCGCTCAACATCACTACACCCAG 1125
Db 331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY 1126 TCCATATAGCTCTAC-----TACGCTTTCAGCAACAGCGCACTTGTGTGGGCC 1179
Db 350 AsnAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY 1180 AACITCATCGAGACCTCGAGAGATCCTGCTCTCCCGTGGTGTCTCCCTCATCTAT 1239

```

```

Db      370 TrpLeuThrGluIleMetAsnAsu-----TyrLysValLeuIleTyrAsn 384
QY      1240 GCGAGCGCGATTACATC-----TCCAC 1263
Db      385 GlyGlnLeuAspIleValAlaAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY      1264 TGGTTCCGGGGTCAGCGGTTTCCTCGGTGGAACTACTCCCAAGCCGCCAG----- 1317
Db      405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY      1318 -----TTCGGAAGC-----GCAGGTACACGCCGCCCTGAAAGTCAACGGC 1356
Db      419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY      1357 GTCGAGTATGGGAACATCCGAGTATGGTAATTTCTCTTCACTCGCGCTCAAGGCA 1416
Db      432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY      1417 GGCCATGAAGTCCCATACACAGCCATCGCTCCCTGCAATTTGTTAACCGGACTATC 1476
Db      447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY      1477 TTC-----GGTGGGAT 1488
Db      467 TyrGlyLysGlyTrpAsp 472

RESULT 13
US-09-707-613-164
; Sequence 164, Application US/0907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hauspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13

```

```

; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-707-613-164

Alignment Scores:
Pred. No.: 1,29e-21      Length: 476
Score: 359.00           Matches: 130
Percent Similarity: 39.51%      Conservative: 62
Best Local Similarity: 26.75%    Mismatches: 178
Query Match: 12.36%            Indels: 116
DB: 11                      Gaps: 21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-907-613-164 (1-476)
QY      139 CCGGGT-----GTCAATCTACTCTGGATAT-----GTGCAACACTCTCCGAGTCC 186
Db      67 ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer 86
QY      187 CATACCTTCTCTGGTCTTGGAGCCAGACATACCCAGAACTGCACCTATCACATTG 246
Db      87 AsnLeuPheTrpPheProAlaGlnProGluAspAlaProValValLeu 106
QY      247 TGGTTGAATGGTGGCCCTGGAGGATCTTTGATCGTCTCTTCGAAGAGTTGGGCCT 306
Db      107 TrpLeuGlnGlyGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro 126
QY      307 TGCCATGTCAATTCGACTTTTGTGATCATCAACCCCTCACTCGTGAACAGAGTCTGCC 366
Db      127 TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProTrpThrThrLeu 146
QY      367 AATTACTATTCTGTCCCGCCAGCCATGGGAGTCGGCTTTTCATATAGTATACGTTGAT 426
Db      147 SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspAspThrHis 166
QY      427 GGGTCCATTAAACCCCTGTAACCTGGGTGCTGTCGCAAAATTCGAGCTTTCGAGGAGTTCAGGGC 486
Db      167 Gly----- 167
QY      487 CGGTACCCACACCATTTGATGCCACTCTGATCGATATACCAATCTTGGCCAGAGGCCGCT 546
Db      168 -----TyrAlaValAsnGluAspValAlaArgAspLeuTyr 180
QY      547 TGGGAGATCTTCGAGAGGATTCCTTAGTGCACACTACCTAGCTTGGACTCTAGGTCAGTCT 606
Db      181 SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrIlyAsn 195
QY      607 AAGGACTTCAGTCTATGCAACGAGAGCTATGGAGGCACATGAGGCTCTGATGTCATCTCAAT 666

```

```

Db      196 AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis 215
QY      667 CAFTTTTACGAGCAAGATGAGAGAAATGCCAACGGTAGTGTTAATGGTTTCACCTTAAT 726
Db      216 LeuIleHisSerLeuAsn-----ProValArgGluValIleAsn 229
QY      727 TTCACACTCTGGGAATATTATACGGCAATCATCGCAGGCGATCCAGGCCCTTACTAC 786
Db      230 LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr 249
QY      787 CCTGAATTCGTGTGGAACAATACCTACGGATATCAAGCTGTCAACGAGACGGCTGTACAAC 846
Db      250 AlaGluPhe-----LeuTyrGluIleGlyLeuLeuAspGluLysGlnLysLys 265
QY      847 TACATGAAGTTTGCACCAAAATGCCAAATGGTTCGCCAGGATTTGATTCACCTGCAAA 906
Db      266 Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg 279
QY      907 CAGACAAC-----CCACCGCATTTAGCTGACTACGCCCTCTCGCGCAGGCCACC 957
Db      280 LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr 299
QY      958 AAC-----AATGTCAGGGCAATGTGAGGG-----CCATACACTACGCTTGTGCTGGT 1005
Db      300 SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu-- 318
QY      1006 CGTGGTGTGTATGATATTCGGCATATATGATGACCGGACTCCGCAAGTTTATACAA 1065
Db      319 -----ArgCysThrGluProGluAspGlnLeuTyrTyrVal 330
QY      1066 AAATTTCTGCCAAGACTCTGTCAATGGAGCTATCGGCGTCAACATCAACTACACCCAG 1125
Db      331 LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe 349
QY      1126 TCAATAATAGCTATAC-----TACGCTTTCCACGACAAACAGCGGACITTTGCTGGCC 1179
Db      350 AsnAspGlyThrIleValGlnLysTyrLeuArgGluAspThrValGlnSerValLysPro 369
QY      1180 ACTTTCATCGAAGCACTCGAGGAGATCCTTGTCTCCCGTGGTGTCTCCCTCATCTAT 1239
Db      370 TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn 384
QY      1240 GCGAGCGCGATATATC-----TCCAAC 1263
Db      385 GlyGlnLeuAspIleIleValAlaLaLeuThrGluArgSerLeuMetGlyMetAsp 404
QY      1264 TGTGTCGGCGTCCAGCGCGTTTCCCTCGCTGCGAATCTCCCAAGCCGCCAG----- 1317
Db      405 TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal 418
QY      1318 -----TTCGGAAGC-----GCAGGTACACGCCCTGAAAGTCAACGCG 1356
Db      419 TrpLysIlePheLysSerAspSerGluValAlaGlyTyr----- 431
QY      1357 GTGAGTATGGGAACACCGCAGTATGTAATTTCTCCTTCACTCGCGCTCATGAGGCA 1416
Db      432 -----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly 446
QY      1417 GSCCATGAATCCCATCTACCGCCATCGCTCCCTGCAATTTGTTAACCGGACTATC 1476
Db      447 GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle 466
QY      1477 TTC-----GGTGGGAT 1488
Db      467 TyrGlyLysGlyTrpAsp 472

```

RESULT 14

```

US-09-907-942-164
; Sequence 164, Application US/0907942
; Publication No. US20030027146A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi

```

```

Alignment Scores:
Pred. No.: 1.29e-21 Length: 476
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 52

```

```

; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Pacini, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,942
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-942-164

```

Best Local Similarity: 26.75% Mismatches: 178			
Query Match: 12.36% Indels: 116			
Db: 11 Gaps: 21			
US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-907-942-164 (1-476)			
QY	139	CGGGT-----GTCAATCTCTACTCTCGATAT-----GTCCAGACCTCTCCCGAGTCC	186
Db	67	ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer	86
QY	187	CATACCTTCTCTGTTCTCGAAGCCACATATACCCAGAACTGCACCTATCATATG	246
Db	87	AsnLeuPhePheTrpPheProAlaGlnLeuGlnProGluAspAlaProValLeu	106
QY	247	TGGTGAATGGTGGCCCTGGAAGCGATTCTTTCATCGGCTCTCTCGAAGAGTTGGCCCT	306
Db	107	TrpLeuGlnGlyProGlySerMetPheGlyLeuPheValGlnHisGlyPro	126
QY	307	TGCCATGTCATATCGACTTTTGATGACTACATCAACCTCTACTCGGTGAACAGAGTCTCC	366
Db	127	TyrValValThrSerAsnMetThrLeuArgAspArgAspPheProThrThrLeu	146
QY	367	AATTTACTATCTCTGCCAGCATATGGAGTGGCTTTTCAATATAGTATAGGTTGAT	426
Db	147	SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspThrHis	166
QY	427	GGGTCCATTACCTCTACTGGGTCGTCGAAANTTCGAGCTTTCGAGGAGTTCAGGGC	486
Db	167	Gly-----	167
QY	487	CGGTACCAACCAATTGATGCCACTCTGATCGATCACTACCAATCTGCCGAGAGCGGCT	546
Db	168	-----TyrAlaValAsnGluAspValAlaArgAspLeuTyr	180
QY	547	TGGAGATCTCTGCAAGATTCTTGTAGTGGACTACCTAGCTTGGAGTGGAGTCT	606
Db	181	SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn	195
QY	607	AAGCACTTCTAGTATGGAGGAGAGTATGGAGGCACTATGCTCTGCATCTTCAAT	666
Db	196	AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis	215
QY	667	CATTTTACGAGCAGATGAGAAATGCCAACGGTAGTGTATATGCTGTCTACGCTTAAT	726
Db	216	LeuIleHisSerLeuAsn-----ProValArgGluValLysIleAsn	229
QY	727	TTCACCTCTCTGGGAATTTATACGGCATCATCGACGAGCGGATCCAGGCCCTTACTAC	786
Db	230	LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyTyr	249
QY	787	CCTGAATTCCTGTGAACATACTACGGTATCAAGGCTCTCAACGAGACCGCTTACAAC	846
Db	250	AlaGluPhe-----LeuTyrGlnIleGlyLeuLeuAspGluLysGlnLysLys	265
QY	847	TACATGAAGTTTGCCAAACCAATGCCAAATGGTTCGAGGATTTGATTTCCACCTGCAA	906
Db	266	Tyr-----PheGlnLysGln-----CysHisGluCysIleGluHisIleArg	279
QY	907	CAGACAAC-----CGCACCGCATTAGCTACACGCTTACGCTCGCCGAGCCACC	957
Db	280	LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr	299
QY	958	AAC-----ATGTCAGGGACAATGTTGAGGG-----CATACTACGCTTTGCTGGT	1005
Db	300	SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnTyrTyrAsnPheLeu---	318
QY	1006	CGTGGTGTATGATATTCGGCATCCATATGATGACCGGACTCCGCCAAGTTATACAC	1065
Db	319	-----ArgCysThrGluProGluAspGlnLeuTyrTyrVal	330
QY	1066	AAATTTTCGCAAGGACTCTGTCTATGACGCTATCGCGTCAACATCACTACACCGAG	1125
Db	331	LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly---AsnGlnThrPhe	349
QY	1126	TCCAATAATGACGTCTAC-----TACGCTTTCCAGCAACAGCCACTTTGTCTGGCCC	1179
Db	350	AsnAspGlyThrIleValGlnLysTyrLeuArgGluAspThrValGlnSerValLysPro	369
QY	1180	AAC TTCATCGAAGACCTCGAGGAGATCCTTGTCTCCCGTGGTCTCTCCCTCATCTAT	1239
Db	370	TrpLeuThrGluIleMetAsnAsn-----TyrLysValLeuIleTyrAsn	384
QY	1240	GCGAGCGCGGATATACATC-----TGC AAC 1263	
Db	385	GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp	404
QY	1264	TGTTTCGCGGCTCAGCGGCTTCCCTCGTCCGCACTACTCCCAAGCCGCCAG-----	1317
Db	405	TrpLysGlySerGln-----GluTyrLysLysAlaGluLysLysVal	418
QY	1318	-----TTCGAGAGC-----GCAGGTACACGCCCTGAAAGTCAACGGC	1356
Db	419	TrpLysIlePheLysSerAspSerGluValAlaGlyTyr-----	431
QY	1357	GTCCAGTATGGGAAACTCGCGAGTATGTTAATTTCTCTCTACTCGCGTCTATGAGGCA	1416
Db	432	-----IleArgGlnAlaGlyAspPheHisGlnValIleIleArgGlyGly	446
QY	1417	GCCATCAAGTCCCATCTACTACCGCCATCCAGCCATCGCTCCCTGCAATTTTACCGGACTATC	1476
Db	447	GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle	466
QY	1477	TTC-----GGTTGGGAT 1488	
Db	467	TyrGlyLysGlyTrpAsp 472	
RESULT 15			
US-09-796-753-40			
; Sequence 40, Application US/09796753			
; Publication No. US20030027998A1			
; GENERAL INFORMATION:			
; APPLICANT: McCarthy, Sean A.			
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF			
; FILE REFERENCE: 7853-227-999			
; CURRENT APPLICATION NUMBER: US/09796,753			
; CURRENT FILING DATE: 2001-03-01			
; PRIOR APPLICATION NUMBER: 09/183,175			
; PRIOR FILING DATE: 1998-10-30			
; PRIOR APPLICATION NUMBER: 09/223,094			
; PRIOR FILING DATE: 1998-12-30			
; PRIOR APPLICATION NUMBER: 09/223,546			
; PRIOR FILING DATE: 1998-12-30			
; PRIOR APPLICATION NUMBER: 09/224,246			
; PRIOR FILING DATE: 1998-12-30			
; PRIOR APPLICATION NUMBER: 09/259,388			
; PRIOR FILING DATE: 1999-02-26			
; PRIOR APPLICATION NUMBER: 60/122,458			
; PRIOR FILING DATE: 1999-03-01			
; PRIOR APPLICATION NUMBER: 09/312,359			
; PRIOR FILING DATE: 1999-05-14			
; PRIOR APPLICATION NUMBER: 09/336,536			
; PRIOR FILING DATE: 1999-06-18			
; PRIOR APPLICATION NUMBER: 09/342,687			
; PRIOR FILING DATE: 1999-06-29			
; PRIOR APPLICATION NUMBER: 05/345,464			
; PRIOR FILING DATE: 1999-06-30			
; PRIOR APPLICATION NUMBER: 09/365,164			
; PRIOR FILING DATE: 1999-07-30			
; PRIOR APPLICATION NUMBER: 09/399,723			
; PRIOR FILING DATE: 1999-09-20			
; PRIOR APPLICATION NUMBER: 09/409,634			
; PRIOR FILING DATE: 1999-09-30			
; PRIOR APPLICATION NUMBER: 09/471,179			
; PRIOR FILING DATE: 1999-12-23			
; PRIOR APPLICATION NUMBER: 09/474,071			
; PRIOR FILING DATE: 1999-12-29			

```

: PRIOR APPLICATION NUMBER: 09/474,072
: PRIOR FILING DATE: 1999-12-29
: PRIOR APPLICATION NUMBER: 09/514,010
: PRIOR FILING DATE: 2000-02-25
: PRIOR APPLICATION NUMBER: 09/516,745
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: 09/572,002
: PRIOR FILING DATE: 2000-05-14
: PRIOR APPLICATION NUMBER: 09/597,993
: PRIOR FILING DATE: 2000-06-19
: PRIOR APPLICATION NUMBER: 09/599,596
: PRIOR FILING DATE: 2000-06-22
: PRIOR APPLICATION NUMBER: 09/630,334
: PRIOR FILING DATE: 2000-07-31
: PRIOR APPLICATION NUMBER: 09/606,565
: PRIOR FILING DATE: 2000-06-29
: PRIOR APPLICATION NUMBER: 09/606,317
: PRIOR FILING DATE: 2000-06-29
: PRIOR APPLICATION NUMBER: 09/665,666
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: 09/677,751
: PRIOR FILING DATE: 2000-09-30
: NUMBER OF SEQ ID NOS: 162
: SEQ ID NO 40
: LENGTH: 476
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-796-753-40

```

Alignment Scores:

Pred. No.:	1,298-21	Length:	476
Score:	359.00	Matches:	130
Percent Similarity:	39.51%	Conservative:	62
Best Local Similarity:	26.75%	Mismatches:	178
Query Match:	12.36%	Indels:	116
DB:	11	Gaps:	21

US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-796-753-40 (1-476)

139	QY	CGGGT-----GTCAATCCTACTCTGGATAT-----	186	QY	GTGACACCTCTCTCCGAGTCC
140	Db	: : : : :	187	Db	: : : : :
67	QY	ProGlyLeuAsnMetLysSerTyrAlaGlyPheLeuThrValAsnLysThrTyrAsnSer	86	QY	: : : : :
187	Db	CATACCTTCTTCTGGTTCTTCGAGCGACAGATAAACCAGAACTGCACCTTACACATG	246	QY	: : : : :
188	Db	: : : : : : : : : : :	247	QY	: : : : :
87	QY	AsnLeuPhePheIrrPhePheProAlaGlnIleGlnProGluAspAlaProValValLeu	106	QY	: : : : :
247	QY	TGGTTGAATGGTGGCCCTCGAAGCGAGTTCTTTTCATCGGTCTCTTCGAAAGAGTTGGCCCT	300	QY	: : : : :
107	Db	TriLeuGlnGlyProGlyGlySerSerMetPheGlyLeuPheValGluHisGlyPro	126	QY	: : : : :
307	QY	TGCCATGTCAATTCGACTTTTGATGACTACATCAACCTCACTCGTGGGAACGAGTCTCC	366	QY	: : : : :
127	Db	TyrValValThrSerAsnMetThrLeuA:gasParGasPheProIrrPheThrThrLeu	146	QY	: : : : :
367	QY	ANITACTATCTGTCCAGCCATTCGGAGTCGGCTTTTCATATAGTACATCGTTGAT	426	QY	: : : : :
147	Db	SerMetLeuTyrIleAspAsnProValGlyThrGlyPheSerPheThrAspAspThrHis	166	QY	: : : : :
427	QY	GGGTCCATTAACCTGTAACTGGGTGGTCGCAAAATTCGAGCTTTTCGAGGATTCAGGCG	486	QY	: : : : :
167	Db	Gly-----	167	QY	: : : : :
487	QY	CGGTACCCCAACCATTGATGCCACTCTGATCGATACCTACCAATCTTCGCGCAGAGCCGCT	546	QY	: : : : :
168	Db	-----TyrAlaValAsnGluAspAspValAlaArgAspLeuTyr	180	QY	: : : : :
547	QY	TGGGAGATCTCGCAAGGATTCCTTAGTGGACTACCTAGCTTCGACTCTAGGTCGAGTCT	606	QY	: : : : :
181	Db	SerAlaLeuIleGlnPhePhe-----GlnIlePheProGluTyrLysAsn	195	QY	: : : : :
607	QY	AAGGACTTCAGTCTATGGACGGAGAGCTATGAGGGGCACTATGGTCCTGCATTCCTTCAT	666	QY	: : : : :

Db	196	AsnAspPheTyrValThrGlyGluSerTyrAlaGlyLysTyrValProAlaIleAlaHis	215
QY	667	CATTTTTCAGCAGCAAGTTCAGAGAATTCACACGGTAGTGTAATGGTGTCTCAGCTTAAT	726
Db	216	LeuIleHisSerLeuAsn	229
QY	727	TTCAACTCTCGGGAAATTATTAAACGGCATCATCGACGAGCGGATCCAGCGCCCTTACTTAC	786
Db	230	LeuAsnGlyIleAlaIleGlyAspGlyTyrSerAspProGluSerIleIleGlyGlyTyr	249
QY	787	CCTGAATTCCGTGTGAACAATACCTACGGTATCAGGCTGTCAAGCAGCAGCGTCTACAAAC	846
Db	250	AlaGluPhe	265
QY	847	TACATGAAGTTGTCACCAACCAATGCCAAATGGTTCACAGGATTTGATTTCACCTGCAAA	906
Db	266	Tyr	279
QY	907	CAGCAACAAC	957
Db	280	LysGlnAsnTrpPheGluAlaPheGluIleLeuAspLysLeuLeuAspGlyAspLeuThr	299
QY	958	AAC	1005
Db	300	SerAspProSerTyrPheGlnAsnValThrGlyCysSerAsnIleTyrAsnPheLeu	318
QY	1006	CGTGGTGTGTATGATATTCGGCATCCATATGATCAGCCGATCCGCCCAAGTTATTACAAC	1065
Db	319		330
QY	1066	AAATTTCTGCCAAAGGACTGTGATCAGAGCTATCGGGGTCAACATCAACATACACCCAG	1125
Db	331	LysPheLeuSerLeuProGluValArgGlnAlaIleHisValGly	349
QY	1126	TCCAATATGACGCTAC	1179
Db	350	AsuAspGlyThrIleValGluLysTyrLeuArgGluAspThrValGlnSerValLysPro	369
QY	1180	AACTTCATCGAAGACCTCGAGGAGATCCTTGCTCCCGGTGGTGTCCCTCATCATAT	1239
Db	370	IrpLeuThrGluIleMetAsnAsn	384
QY	1240	GGCGACGCCATTACATC	1263
Db	385	GlyGlnLeuAspIleIleValAlaAlaLeuThrGluArgSerLeuMetGlyMetAsp	404
QY	1264	TGGTTCCGGCGGTACGGCGGTTTCCTCGCTGCGAACTACGCCAAGCCGCCAG	1317
Db	405	TrpLysGlySerGln	418
QY	1318		431
Db	419	TrpLysIlePheLysSerAspSerGluValAlaIleGlyTyr	431
QY	1357	GTCCAGTATGGGAAACTCCGAGTAGTGTAAATTCCTTCTACTCGGTCTATGAGGCA	1416
Db	432		446
QY	1417	GGCCATGAATCCCATCTACCAACGCCCATCGCCTCCCTGCAATTTGTTTAACCGGACTATC	1476
Db	447	GlyHisIleLeuProTyrAspGlnProLeuArgAlaPheAspMetIleAsnArgPheIle	466
QY	1477	TTC	1488
Db	467	TyrGlyLysGlyTrpAsp	472

Search completed: September 16, 2003, 18:56:17
Job time : 211 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:39:03 ; Search time 47 Seconds

(without alignments)
6580.393 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662

Perfect score: 2904

Sequence: 1 cttccagggaagtacaccggc.....gcattgctcagtggtggtatg 1608

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DRV=xlp
-Q=/cgn2.1/USPTO.spool_p/US09712338/runat_16092003_144326_14378/app_query.fasta_1.1799
-DB=PIR_76 -OPMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338_@cgn_1.1.57_@runat_16092003_144326_14378 -NCPD=6 -ICPD=3
-NO_MMAPP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG
-DRV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	30.8	423	2	S38953 carboxypeptidase D
2	746.5	25.7	460	1	S51516 serine-type carbox
3	652.5	22.5	1002	2	T43236 carboxypeptidase C
4	636	21.9	508	1	S45008 probable carboxype
5	601	20.7	552	2	JC7666 serine-type carbox
6	596	20.5	542	1	JC1380 carboxypeptidase C
7	594.5	20.5	523	1	S61713 carboxypeptidase C
8	527.5	18.2	532	1	CEBYU probable carboxype
9	473	16.3	429	1	T03607 carboxypeptidase C
10	452	15.6	500	1	S22530 carboxypeptidase-1
11	439	15.1	510	2	T48977 carboxypeptidase C
12	425.5	14.7	411	1	A35275 carboxypeptidase C
13	413.5	14.2	499	1	A29412 carboxypeptidase C
14	413.5	14.2	2105	2	T18968 probable serine-ty

15	383	13.2	482	2	T49079 serine-type carbox
16	363	12.5	480	2	T50511 serine-type carbox
17	362.5	12.5	487	2	T49080 serine-type carbox
18	358	12.3	470	2	B96637 hypothetical prote
19	352.5	12.1	470	2	T16606 probable serine ca
20	351	12.1	482	2	A43828 probable serine ca
21	349	12.0	474	2	A35732 protective protei
22	348	12.0	452	2	H84772 probable serine ca
23	344.5	11.9	516	2	S44191 carboxypeptidase D
24	344	11.8	465	2	G86244 Serine carboxypept
25	344	11.8	501	2	T49081 probable serine ca
26	340	11.7	512	2	T33463 probable serine ca
27	337.5	11.6	425	2	F85360 SERINE CARBOXYPEPT
28	333.5	11.5	510	2	T39601 probable serine ca
29	333	11.5	487	2	B8472 probable serine ca
30	332.5	11.4	480	2	A31589 carboxypeptidase C
31	332	11.4	465	2	B85358 SERINE CARBOXYPEPT
32	332	11.4	469	2	T16316 probable serine ca
33	330.5	11.4	474	2	D84631 hypothetical prote
34	330	11.4	1203	2	T21275 hypothetical serine ca
35	327.5	11.3	2338	2	T25810 probable serine ca
36	327.5	11.3	479	2	G96501 KEX1 protein precu
37	326.5	11.2	425	2	B84631 probable serine ca
38	326.5	11.2	454	2	T20829 probable serine ca
39	321.5	11.1	729	2	A29651 serin carboxypepti
40	320.5	11.0	456	2	H86406 probable serine ca
41	319	11.0	502	2	T49188 probable serine ca
42	318	11.0	468	2	D84503 hypothetical prote
43	316	10.9	574	2	T16230 carboxypeptidase D
44	314.5	10.8	476	1	T05701 carboxypeptidase C
45	312	10.7	510	1	S43516

ALIGNMENTS

RESULT 1

S38953
carboxypeptidase D (EC 3.4.16.6) - Penicillium janthinellum
N: Alternate names: carboxypeptidase S1
C: Species: Penicillium janthinellum
C: Date: 18-Feb-1994 #sequence_revision 19-Apr-1996 #text_change 31-Mar-2000
C: Accession: S38953
R: Svendsen, I.; Hofmann, T.; Endrizzi, J.; Remington, S.J.; Breddam, K.
FEBS Lett. 333, 39-43, 1993
A: Title: The primary structure of carboxypeptidase S1 from Penicillium janthinellum.
A: Reference number: S38953; MUID:94039747; PMID:8224168
A: Accession: S38953
A: Status: preliminary
A: Molecule type: protein
A: Residues: 1-423 <SVE>
C: Superfamily: serine carboxypeptidase
C: Keywords: hydrolase; serine carboxypeptidase

Alignment Scores:
Pred. No.: 9.45e-53 Length: 423
Score: 894.00 Matches: 186
Percent Similarity: 54.31% Conservative: 66
Best Local Similarity: 40.09% Mismatches: 160
Query Match: 30.79% Indels: 52
DB: 2 Gaps: 8

US-09-712-338-1_COPY_55_1662 (1-1608) x S38953 (1-423)

QY 121 GCGCTGCGAGACTACCGGGTGTCAAATCCTACTCTGGATATTCGACACCTCTCCC 180
|||||
Db 6 GlylleCysGluThrThrProGlyValAsnGlnItyrSerGlyItyrLeuSerValGlySer 25
QY 181 GAGTCCCATCCTCTCTCTGGTCTCTCGAAGCCAGACATACCCAGAACTCCACCTATC 240
|||
Db 26 AsnMetAsnMetTrpPheTrpPheGluAlaArgAsnAsnProGlnGlnAlaProLeu 45
QY 241 ACATGTGTGATGATGGTGGCCCTCGAAGCGAATCTTTGATGGGTCTCTCGAAGATTG 300
|||

Db 46 AlaAlaIrrPpAsnGlyProGlyCysSerSerMetIleGlyLeuPheGlnGluAsn 65
QY 301 GGCCCTTCCCAT---GTCAATTCGACTTTTGTATGACTACATCAACCCCTCACTCGTGGAC 357
Db 66 GlyProCysHisPheValAsnGlyAspSerThrProSerLeuAsnGluAsnSerIrrPpn 85
QY 358 GAGGTCCTCAATTTACTATCTCTGCCAGCCATGGGAGTGGCGCTTTTCATATAGTAT 417
Db 86 AsnTyrAlaAsnMetIleTyrIleAspGlnProIleGlyValGlyPheSerTyrGly--- 104
QY 418 ACGGTTGATGGTCCATTAAACCTCTGAACCTGGGGTCTGCGAAATTCGAGCTTTCAGGA 477
Db 104 ----- 104
QY 478 GTTCAGGGCGGTACCCACCAATTGATGCCACTCTGATCGATACCTACCAATCTGCCGA 537
Db 105 -----ThrAspAspValThr-----SerThrValThrAlaAla 115
QY 538 GAGGCGGCTTGGGAGATCTCTGCAAGGATTCCTTAGTGGACTACCTAGCTGGACTTAGG 597
Db 116 ProTyrValIrrPpnAsnLeuGlnAlaPheTyrAlaGlnArgProGlnTyrGluSerArg 135
QY 598 GTCCAGCTPAAGACTTCAGTCTATCGACGGAGAGCTATGGAGGGACATAGTCTCGCA 657
Db 136 -----AspPheAlaIlePheThrGluSerTyrGlyGlyHisTyrGlyProGlu 151
QY 658 TTCTTCAATCATTTTACGAGCAGATGAGAGAAATGCCAAGTAGTGTATATGCTGT 717
Db 152 PheAlaSerTyrIleGlnGlnAsnAlaAlaIleLysAlaGlySerValThrGlyGln 171
QY 718 CAGCTTAATTAACCTCTGGGAATTATTAAAGCGCATCATCGAGGCGATCCAGGCC 777
Db 172 AsnValAsnIleValAlaLeuGlyValAsnAsnGlyTrpIleAspSerThrIleGlnGlu 191
QY 778 CCTACTACCTGAATTCGCTGGAACAATACCTACGGTATCAAGGCTGTCAAGAGAAC 837
Db 192 LysAlaTyrIleAspPheSerTyrAsnAsnSerTyrGlnGlnIleIleAspSerThr 211
QY 838 GTCTAACAATACATGAAGTTTGGCCACCAATGCCAAATGGTGGCCAGGATTTGATTCC 897
Db 212 ArgAspSerLeuLeuAspAlaTyrAsn-----AsnGlnCysLeuProAlaLeuGln 228
QY 898 ACCTGCAACAGCAACACCGCAGCTAGCTACCTCCCTCTCGCCGAGGCAAC 957
Db 229 GlnCysSerGlnSerGlySerThr-----SerAspCysThrAsnAlaAsp 243
QY 958 AACATGTCAGGACAAATGTTAGGGCCATATACACCTTTCGTGCTGCTGTAT 1017
Db 244 SerValCysTyrGlnAsnIleGluGlyProIleSerSerGlyAspPheAspValTyr 263
QY 1018 GATATTGGCATCCATATGATGACCGACCTCCGCAAGTTATTACAAATTTCTGGCA 1077
Db 264 AspIleArgGluProSerAsnAspProTyrProIlySerThrTyrLeuSer 283
QY 1078 AAGCACTCTCATGTCAGGCTATCGCGCTCAACATCAACATACCCAGTCCATPAATGAC 1137
Db 284 AspProThrValValIysAlaIleGlyAlaArgThrAsnTyrGlnGluCysProAsnGly 303
QY 1138 GTCTACTACGCTTTCAGCAACAGGAGGACTTTCTCGGCCCACTTCATCGAGAGCTC 1197
Db 304 ProTyrAsnLysPheAlaSerThrGlyAspAsnPro---ArgSerPheLeuSerThrLeu 322
QY 1198 GAGGAGATCTGTCTCCCGTCTCCCTCATCTATCGGCGAGCGGATTAAC 1257
Db 323 SerSerValValGlnSerGlyIleAsnValLeuValIrrPpnAlaGlyAspAlaAspTrpIle 342
QY 1258 TGCAACTGTTTCGGGCTAGGCGCTTCCCTCGCTCGCACTACTCCCAAGCCGCCAG 1317
Db 343 CysAsnIrrPpnLeuGlyAsnTyrGluValAlaAsnAlaValAspPheProGlyAsnAlaGln 362
QY 1318 TTCGGAACGCGGATACACGCCCTCGAAGTCAACGGCGTCGATATGGGAACTCGC 1377
Db 363 PheSerAlaLeuAspLeuAlaProTyrThrValAsnGlyValGluLysGlyGlnPheLys 382

QY 1378 GAGTATGGTAATTTCTCTCTACTCCGCTCTATGAGGAGGAGCCATCAAGTCCCATCTAC 1437
Db 383 ThrValAspAsnPheSerPheLeuLysValIrrGlyAlaGlyHisGluValProTyrTyr 402
QY 1438 CAGCCCATCGCTCCCTCGCAATTTTAAACCGGACTATCTCGTGGGATATCGCAGAG 1497
Db 403 GlnProAspThrAlaLeuGlnAlaPheLysGlnIleIle----- 415
QY 1498 GCCCAGCAAGAG 1509
Db 416 ---GlnLysLys 418
RESULT 2
S51516
serine-type carboxypeptidase (EC 3.4.16.-) Z precursor - Absidia zycha
C:Species: Absidia zycha
C:Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text_change 16-Jun-2000
C:Accession: S51516; S78013; S78014
R:Lee, B.R.; Takeuchi, M.; Kobayashi, Y.
Curr. Genet. 27, 159-165, 1995
A:Title: Molecular cloning and sequence analysis of the sopz gene encoding the serine
A:Reference number: S51516; MUID:95308538; PMID:7788719
A:Accession: S51516
A:Molecule type: DNA
A:Residues: 1-460 <LEE>
A:Cross-references: EMBL:D16519; NID:g556466; PIDN:BA03966.1; PID:g995456
A:Experimental source: strain NRIC 1199
A:Accession: S78013
A:Molecule type: protein
A:Residues: 52-62; 90-99; 367-381 <LEB>
A:Accession: S78014
A:Molecule type: mRNA
A:Residues: 18-460 <LES>
C:Genetics:
A:Gene: sopz
A:Introns: 64/3; 95/1; 107/2; 138/3; 234/2; 246/1; 270/1; 343/3; 353/2; 384/3; 437/3
C:Superfamily: serine carboxypeptidase
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-51/Domain: propeptide #status predicted <PRE>
F:52-460/Product: serine-type carboxypeptidase #status experimental <MAT>
F:40,46,119/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:186/Active site: Ser #status predicted
Alignment Scores:
Pred. No.: 9.44e-43 Length: 460
Score: 746.50 Matches: 175
Percent Similarity: 50.31% Conservative: 71
Best Local Similarity: 35.79% Mismatches: 182
Query Match: 25.71% Indels: 61
DB: 1 Gaps: 14
US-09-712-338-1_copy_55_1662 (1-1608) x S51516 (1-460)
QY 16 CCGCGCTCGCTCGTAGAGACAGTACCCCAAGACCCCGGGTCAAGACTCTTACA 75
Db 29 ProMetSerHisValLeuArgGlnAspGlyAsnAspThr-----Ser 43
QY 76 ACCGCAACAATGTCACCATCCGTTACAGGAACCCGGGAGAGGGCGTCTGGAGACT 135
Db 44 SerGlyAsnThrThrGlnLeuLysTyrThrSerPro-----LysLeuCysAsp--- 59
QY 136 ACCCGGGTGTCAATCTCTACTCTGGATATGTCGACACTCTCGGAGTCCCATCTTC 195
Db 60 ---ProAspValLysGlnTyrSerGlyTyrLeuAspAlaAlaAsnAspGluHisTyrPhe 78
QY 196 TTCGGTCTTCGAGACCCAGACATACCCAGAACTGCACCTATCACATTTGGTTGAAT 255
Db 79 PheTrrPhePheGluSerLysAsnAspProLysAsnAspProLeuThrIrrPpnLeuAsn 98
QY 256 GGTGGCCCTCGAAGCGATTTCTTGTATCGGTCTCTTCGAGAGACTGGCCCTTGCCTATGTC 315

Db 99 GlyGlyProGlyCysSerSerLeuIleGlyLeuTrpGluGluLeuGlyProCysGlnGln 118
QY 316 AATTCGACTTTTATGATCAATCAACCCCTACCTCGTGGACGAGGCTCCAAATTTACTA 375
Db 119 AsnGlySer-----AlaAsnProHisSerTrpHisSerSerAsnMetLeu 134
QY 376 TTCCTGTCCGAGCATTTGGAGTGGCGTTTTCATATAGTGATACGGTTGATGGGTCCAIT 435
Db 135 PhePheAspGlnProAspGlyValGlyPheSerTyr----- 146
QY 436 AACCTCTGAACCTGGGTCTCGAATAATTCAGAGCTTCAGAGATTCAGGCGCGGTACCCA 495
Db 147 -----GlyLysGln----- 149
QY 496 ACCAATGATGCCACTCTGATGATACCAATCTTCGCCGACAGGCGCTTGGAGATC 555
Db 150 -----ThrValSerThrThrGluAspAlaGluArgAlaTrpThrPhe 164
QY 556 CTCGAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGTGTCAGTCTAAGGACTTC 615
Db 165 LeuGlnAlaPheTyrGluThrPhePro-----GlnTyrSerLysLeuAspVal 180
QY 616 AGTCTATGGACGAGACTATGAGGCGCACTATGGTCTGCATCTTCAATCATCTTTIAC 675
Db 181 HisTyrPheGlyGluSerTyrGlyGlyHisTyrIleProGlyPheAlaSerHisValVal 200
QY 676 GAGCAGAATAGAGAATTGCCACGGTAGTGTAAATGGTGTTCAGCTTAATTTCAACTCT 735
Db 201 AspMetAsnLysLysValGlnSerGlyGluGlyLysGlyValValProLeuLysSer 220
QY 736 CTGGGAATTATTAAACGGCATCATCGACGAGGCGATCCAGCCCTTACTACCTGATTC 795
Db 221 IleGlyValGlyAsnGlyPheIleAspAlaValIleGlnTyrLysSerTyrProLysMet 240
QY 796 GCTGTGAACAATACCTACCGATCAAGGCTGTCAACGAGACCGCTCAACTACATGAAG 855
Db 241 ThrCysHisSerThrTyrProAlaValLeuSerGluGlu-----GluCysAspLys 257
QY 856 TTTGCCAACCAATGCCAATGTGTCGAGGATTTGATTTCCACCTGCAACAGACAAC 915
Db 258 MetGlnGlnIleTyrGluAsnAspCysLysProAlaAlaGluGlnCysAlaGluSerAsp 277
QY 916 CGCACCGCATTAGCTGACTACCGCCCTCGCGCGAAGCCACCACCACTGTGCGGGACAAT 975
Db 278 -----GluAspSerAspCysValAsnAlaAsnGlnGlnCys---GlyGln 291
QY 976 GTTAGGGGCGCATACCGCTTTCGTGCTGCTGCTGTGTATGATATTCGGCAICCAT 1035
Db 292 IleGluGly---IleTyrAlaGlnSerGlyTyrSerPheTyrAspIleArgGlnGlnGly 310
QY 1036 GATGACCGGACCTCGGCAAGTTATTACAACAATTTCTGGCAAGGACTGTGTCATGGAC 1095
Db 311 AspAsp---ThrProHisProPheValAspGlu---LeuAsnLysAlaSerValIleLys 328
QY 1096 GCTATCGGCGTCAACATCACTACCCAGTCCAAATATAGCTCTACTACGTTTCCAG 1155
Db 329 GluValGlyAlaArgGlyHisPheSerMetCysSerAspSerValGlyThrAlaPheAla 348
QY 1156 CAACAGGCGACTTTGCTGCGCCCACTTCATCGAAGACCTCGAGAGATCCTTGTCTCTC 1215
Db 349 GlnThrGlyAspGlyAla---ArgSerTyrIleProAlaValGluLysLeuLysGlu 367
QY 1216 CCGGTGCGTGTCTCCCTCATCTATGGCGACGCGATTACATCTGCACTGGTGTGGCGGT 1275
Db 368 GlyIleProValLeuIleTyrValGlyAspAlaAspValIleCysAsnTrpIyrGlyAsn 387
QY 1276 CAGGCGGTTTCCTCGCTCGCAACTACTCCCAAGCGCCGACGTTCGAAGCGCAGGGTAC 1335
Db 388 LeuAspValAlaAspSerLeuLysTyrAspGlySerAlaPheSerLysThrLysLeu 407
QY 1336 ACGCCCTGAAGTCAACGGCGTCCAGTATGGGAAACTCGCGAGTATGTTCTCTCC 1395
Db 408 GluAlaTrpLysAlaAspGlyLysGluValGlnPheArgSerAlaAspLysLeuThr 427

QY 1396 TTCACTCCGCTCTATGAGGCGCATCAAGTCCCATACTACCCAGCCCATGCCCTCCCTG 1455
Db 428 PheValArgValTyrGluAlaGlyHisGluValProMetTyrGlnProGluAlaLeu 447
QY 1456 CAATTGTTTTAACCGGACTATCTCGGI 1482
Db 448 SerMetPheGlnThrTrpIleSerGly 456
RESULT 3
T43236
carboxypeptidase C (EC 3.4.16.5) precursor [validated] - fission yeast (Schizosacchar
N:Alternate names: carboxypeptidase Y
C:Species: Schizosaccharomyces pombe
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 08-Sep-2000
C:Accession: W43236; T37997
R:Tabuchi, M.; Iwahara, O.; Ohtani, Y.; Ohuchi, N.; Sakurai, J.; Morita, T.; Iwahara
J. Bacteriol. 179, 4179-4189, 1997
A:Title: Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport,
A:Reference number: Z2357; MUID:97352672; PMID:9209031
A:Accession: T43236
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1002 <TAB>
A:Cross-references: EMBL:D86560; NID:g3046860; PIDN:BAA25568.1; PID:g3046861
R:Oliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, June 1997
A:Reference number: Z21760
A:Accession: T37997
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1002 <OLI>
A:Cross-references: EMBL:Z97209; PIDN:CA10121.1; GSPDB:GN00066; SPDB:SPAC19G12.10C
A:Experimental source: strain 972h(-); cosmid c19G12
C:Comment: The mature carboxypeptidase exists as a heterodimer held together by a dis
C:Genetics:
A:Gene: cpy1; SPAC19G12.10C
A:Map position: 1
C:Superfamily: serine carboxypeptidase
C:Keywords: disulfide bond; glycoprotein; heterodimer; hydrolase; serine carboxypepti
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-521/Domain: propeptide #status predicted <PRO>
F:200-332/Region: 13-residue repeats (H-H-[EK]-P-G-E-H-M-P-P-P-[MF])
F:344-425/Region: 9-residue repeats (E-H-H-[RQ]-G-G-[KE]-[DL]-K)
F:522-1002/Product: carboxypeptidase C #status experimental <MAT>
F:627-880/Disulfide bonds: #status predicted
F:659/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:715,921,978/Active site: Ser, Asp, His #status predicted
Alignment Scores:
Pred. No.: 2,62e-36 Length: 1002
Score: 652.50 Matches: 165
Percent Similarity: 47.34% Conservative: 66
Best Local Similarity: 33.81% Mismatches: 174
Query Match: 22.47% Indels: 83
DB: 2 Gaps: 15
US-09-712-338-1_copy_55_1662 (1-1608) x T43236 (1-1002)
QY 91 ACCATCCGTACAGGAACCCGGGCGAGGGCGTCTGGAGACTACCCGGGTGCAAA 150
Db 570 ThrLeuArgValLysAspSerLysProGluSerLeuGlyIleAspThr-----ValLys 587
QY 151 TCCTACTCTGGATGTGACACCTCTCCCGAGTCCCATACCTCTCTGTGTTCTTCGAA 210
Db 588 GlnTyrThrGlyTyrLeuAspValGluAspArgHisLeuPheThrPheGlu 607
QY 211 GCCAGACATAACCCAGAACTGCACCTATCACATTGTGTTGAATGGTGGCCCTGGAAGC 270
Db 608 SerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsnGlyGlyProGlyCys 627
QY 271 GATTCTTTCATCGTCTCTTCGAGAGTGGCCCTTCGCATGTCAT---TCGACTTTT 327
Db 327

Db 628 SerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIleAsnIleGluThrLeu 647
QY 328 GATGACTACATCAACCCCTACCTCGTGGAGAGGCTCCAAATTACTATTCTGTCGCCAG 387
Db 648 LysProGluTyrAsnProHisSerIleAsnSerValIlePheLeuAspGln 667
QY 388 CCATTGGGAGTGGCTTTTCATATAGTACGGTGTGATGGGTCCATTAAACCTGTAAC 447
Db 668 ProIleAsnThrGlyPheSerAsnGlyAsp-----AspSerValLeuAspThrValThr 685
QY 448 GGGTGGTGGAAAATTTCGAGCTTTCGAGGAGTTCAGGGCGGTACCCAAACATTGATGCC 507
Db 685 ----- 685
QY 508 ACTCTGATCATCTACCATCTTTCGCGAGAGCGCTTGGGAGATCCTGCAAGGATTC 567
Db 686 -----AlaGlyLysAspValTyrAlaPheLeuAsnLeuPhe 697
QY 568 CTTAGTGGACTACGTAGCTTGGACTCTAGGGTGCAGCTTAAGGACTCTAGTCTATGGACG 627
Db 698 PheAlaLysPhePro-----GlnTyrAlaHisLeuAspPheHisIleAlaGly 713
QY 628 GAGAGCTGAGGGGAGCTATGCTGCTCATCTTCAATCATTTTACGAGCAGATGAS 687
Db 714 GluSerTyrAlaGlyHisTyrIleProGlnPheAlaLysGluIleMetGluHisAsnGln 733
QY 688 AGA-----ATTCGCAACGGT---AGTGTAAATGGTTCAGCTTAATTTCAAC 732
Db 734 GlyAlaAsnPhePheValAlaSerGlyTyrGluMetGluLysGlnTyrIleAsnLeuLys 753
QY 733 TCCTCGGAATTATTACGATCATCGAGCGATCCAGCGATCCGCTTACTACCCCTGAA 792
Db 754 SerValLeuIleGlyAsnGlyLeuThrAspProLeuValGlnTyrPheTyrGlyLys 773
QY 793 TTCGCTGTGAACATCTACGGT-----ATCAAG 822
Db 774 MetAlaCysGluSerProTyrGlyProIleMetSerGlnGluLysAspArgIleThr 793
QY 823 GCTGTCAACGAGCGCTCACTACATCATCAAGTTTCCCAACCAATGCCAAATGGTTCG 882
Db 794 GlyAlaTyrAspThr-----Cys 799
QY 883 CAGGATTTGATTTCCACCTGCAACACAGACACAAACCGCACCGCATAGCTAGCGCCTC 942
Db 800 AlaLysLeuIleThrGlyCysTyrGlnThrGlyPheThr-----ProVal 814
QY 943 TGCCCGCAACCCACCAACATCTCGAGGACAAATGTCAGGGCCATCTAGCGCTTTGCT 1002
Db 815 CysIleGlyAlaSerLeuTyrCysAsnAlaMetIleGlyPro---PheIleLysThr 833
QY 1003 GGTGCTGTGTATGATATTCGGCATCCATCATGATGACCGGACTCCGCCAAGTAT--- 1059
Db 834 GlyLeuAsnIleTyrAspIleArgGluLysArgAspGlnGluHisLeuLysTyrPro 853
QY 1060 -----TACACAAATTTCTGGCAAGGACTCTGTATGAGCGGTATCGGGCTC 1107
Db 854 GluThrGlyAlaIleGluSerTyrLeuAsnGlnGluPheValGlnGluAlaLeuGlyVal 873
QY 1108 AACATCACTACACCCAGTCCAAATATGAGTCTACTACCGTTTCCACCAACAGCGGAC 1167
Db 874 GluTyrAspTyrLysGlyCysAsnThrGluValAsnIleGlyPheLeuPheLysGlyAsp 893
QY 1168 TTTCGTGGCCCACTTCATCGAAGACTCCGAGAGATCCCTP-----GCTCTCCCGGTG 1221
Db 894 TrpMetArgLysThrPheArgAspAspValThrAlaIleLeuGluAlaGlyLeuProVal 913
QY 1222 CGTGTCTCCCTCATCTATP---GGCGACCGCGATTCATCTCAACTGCTTCGGCGGTACG 1278
Db 914 -----LeuIleTyrAlaGlyAspAlaAspTyrIleCysAsnTyrMetGlyAsnGlu 930
QY 1279 GCCGTTTCCCTCGCTGGCACTACTCCCAACCGCCCGAGTCCGAGCGGAGGTACAGG 1338
Db 931 AlaTrpThrAspAlaLeuGluTrpAlaGlyGlnArgGluPheTyrGluAlaGluLeuLys 950

RESULT 4

S46008
probable carboxypeptidase C (EC 3.4.16.5) YBR139w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR1015; hypothetical protein YBR139w
C:Species: Saccharomyces cerevisiae
C>Date: 26-Aug-1994 #sequence_revision 05-Jan-1996 #text_change 19-Apr-2002
C:Accession: S46008; S46581
R:Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45995
A:Accession: S46008
A:Molecule type: DNA
A:Residues: 1-508 <BEC>
A:CROSS-references: EMBL:Z36008; NID:g536435; PIDN:CAA85097.1; PID:g536436; GSPDB:GNO
R:Becam, A.M.; Cullin, C.; Grzybowska, E.; Lacroute, F.; Nasr, F.; Ozier-Kalogeropoul
Yeast 10(Suppl.A), S1-S11, 1994
A:Title: The sequence of 29.7kb from the right arm of chromosome II reveals 13 comple
A:Reference number: S46369; MUID:94378/17; PMID:8091856
A:Molecule type: DNA
A:Residues: 1-508 <BE2>
A:CROSS-references: EMBL:X75891; NID:g496856; PIDN:CAA53497.1; PID:g496869
A:Experimental source: strain S288C
C:Genetics:
A:CROSS-references: SGD:S0000343
A:Map position: 2R
A:Note: MIPS:YBR139w
C:Superfamily: serine carboxypeptidase
C:Keywords: hydrolase; serine carboxypeptidase; transmembrane protein
F:6-22/Domain: transmembrane #status predicted <MM>
F:219,415,474/Active site: Ser, Asp, His #status predicted
Alignment Scores:
Pred. No.: 2,95e-35 Length: 508
Score: 636.00 Matches: 167
Percent Similarity: 46.29% Conservative: 70
Best Local Similarity: 32.62% Mismatches: 197
Query Match: 21.90% Indels: 78
Gaps: 14
DB:
US-09-712-338-1_COPY_55_1662 (1-1608) x S46008 (1-508)
QY 1 CTTCCAGGAAGTACACCGCGCTCGGTGAGAGACAGCTA-----CCCAAG 48
Db 43 LeuProGlnAsnThrGlnGlnThrLeuLysLeuAspArgLeuAsnHisAspProLeu 62
QY 49 AACCCACCGGGTCAAGACTCTACACCCCAACATGTCACCATCCGTTACAGGAA 108
Db 63 PheThrThrPheIleSerSerValAspThrAspTyrSerLeuArgLeuArgThrValAsp 82
QY 109 CCGGGGCGAGGGCGTCTGCGAGACTACCCGGGTGTCAAATCCTACTCTGGATATCTC 168
Db 83 ProSerLysLeuGlyIle-----AspThrValLysGluTrpSerGlyTyrMet 98
QY 169 GACACCTCTCCGAGTCCCATACCTTCTTCGTTCTTCGAGCCAGACATACCCAGNA 228
Db 99 AspTyrLysAspSerLysHisPheTyrTrpPheGluSerArgAsnAspProAla 118
QY 229 ACTGCACCTATCACATTTGGTGTGAATGCTGSCCCTGGAGAGCGATTCTTTGATCGCTCTC 288

```

Db      119 AsnAspProIleLeuThrPheSerAsnGluLeuGluThrPheLeuAsnGlyCysSerPheThrGlyLeu 138
QY      289 TTCAAGAGTTGGGCGCTTCCCATTCATTCGACITTTGATGACTACATCAACCCAC 348
Db      139 LeuPheGluLeuGlyProSerSerIleGlyAlaAspMetLysProIleHisAsnProThr 158
QY      349 TCTGGACAGAGTCTCCAAATTTACTATTCTCTCCAGCCATTCGGAGTCGGTTTCA 408
Db      159 SerTrpAsnAsnAlaSerMetIlePheLeuGluGlnProLeuGlyValGlyPheSer 178
QY      409 TATAGTGATACGGTTCATGAGTCCATTAACCCCTGTAACCTGGGTGTCGAAAATTCGAGC 468
Db      179 TyrGlyAsp----- 181
QY      469 TTTCAGGAGTTTCAGGCGCGGTACCAACCATTTGATGCCACTGTGALCGATACCAAI 528
Db      182 -----GluLysValSerThrLys 188
QY      529 CTTCGCCAGAGCCGCTTGGGAGATCCTGCAAGGATTCCTAGTGGACTACCTAGCTG 588
Db      189 LeuAlaGlyLysAspAlaTyrIlePheLeuGluLeuPheGluAlaPheProHisLeu 208
QY      589 GACTCTAGGCTCAGCTAAGACTTCAGTCTATGACGAGAGAGCTATGAGGCGCTAT 648
Db      209 -----ArgSerAsnAspPheHisIleAlaGlyGluSerTyrAlaGlyHisTyr 224
QY      649 GGTCTGTGCAATTCATCATTTTACGAGCAGAAAT---GAGAGAATTCGCAACGGTAGT 705
Db      225 IleProGlnIleAlaHisGluIleValValLysAsnProGluArg----- 239
QY      706 GTTATGGTGTTCAGCTTAATTCACCTCTCGGGAAATTTAAACGGCATCAFCGACGAG 765
Db      240 -----ThrPheAsnLeuThrSerValMetIleGlyAsnGlyIleThrAspPro 255
QY      766 GCGATCCAGGCCCTTACTACCTGATTCGCTGTG---AACAAATACACGAGTATCAAG 822
Db      256 LeuIleGlnAlaAspTyrIleGluProMetAlaCysGlyLysGlyTyrHisProVal 275
QY      823 GCTGTACAGCAGCGCTGTACAACTACATGAAGTTTGGCAACCAATGCCAATGGTTC 882
Db      276 LeuSerSerGluGluCysGluLysMetSerLysAlaAlaGlyArg----- 290
QY      883 CAGGATTTGATTTCCACCTGCAACAGACAAACCGACCGCATTAAGTACACGCCCTC 942
Db      291 -----CysArgArgLeuAsnLysLeuCysTyrAlaSerLysSerSer 304
QY      943 -----TGCGCCGAAGCCACCAACATGTCAGGAGCAATGTTCAGGGGCCATACAGGC 996
Db      305 LeuProCysIleValAlaThrAlaTyrCysAspSerAlaLeuLeuGluProTyrIleAsn 324
QY      997 TTGCTGCTGTGGTGTATGATATTCGGCATCCATATGATGACCCG----- 1044
Db      325 ---ThrGlyLeuAsnValTyrAspIleArgGlyProCysGluAspAsnSerThrAspGly 343
QY      1045 -----ACTCGCCCAAGTTATACAAACAAATTTCTGGCAAGGACTCTGTCTATGAC 1095
Db      344 MetCysTyrThrGlyLeuArgTyrValAspGlnTyrMetAsnPheProGluValGlnGlu 363
QY      1096 GCATCGGCGTCACATC---AATACACCCAGTCCATAATAGCTGTACTACCTTTC 1152
Db      364 ThrLeuGlySerAspValHisAsnTyrSerGlyCysAspAsnAspValPheThrGlyPhe 383
QY      1153 CAGCAACAGCGGCTTGTCTGCGCCCACTTCATCGAAGACCTCGAGGAGCTCTTGC 1212
Db      384 LeuPheThrGlyAspGlySerLysPro---PheGlnGlnTyrIleAlaGluLeuAsn 402
QY      1213 CTCCCGGTGCTGTCTCCCTCATCTATGGCAGCGCGGATACATCTGCAACTGGTTCGGC 1272
Db      403 HisAsnIleProValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGly 422
QY      1273 GGTCAAGCGGCTTCCCTCGCTGCGCAACTACTCCCAAGCGCGGCTTCCGAGCGAGGG 1332

```

```

Db      423 AsnHisAlaTrpSerAsnGluLeuGluThrPheLeuAsnLysArgArgIleGlnArgMet 442
QY      1333 TACACGCC---CTGAAAGTCAAGCGCGTCGAGTATGGGAAACTCGCGAGTATGGI 1386
Db      443 LeuArgProTrpValSerLysGluThrGlyGluGluLeuGlyGlnValLysAsnIleGly 462
QY      1387 AATTTCTTCCTCCTCCTGCTATGAGCAGCGCCATGAAGTCCCATCTACACGCCCATC 1446
Db      463 ProPheThrPheLeuArgIleTyrAspAlaGlyHisMetValProTyrAspGlnProGlu 482
QY      1447 GCCCTCCCTGCAATGTGTTAACCGGACTATCTTCGGT 1482
Db      483 AlaSerLeuGluMetValAsnSerTrpIleSerGly 494

RESULT 5
JC7666
serine-type carboxypeptidase homolog precursor - Emericella nidulans
N:Alternate names: carboxypeptidase Y homolog
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 03-Aug-2001 #sequence:revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: JC7666
R:Onsumi, K.; Matsuda, Y.; Nakajima, H.; Kitamoto, K.
Biosci. Biotechnol. Biochem. 65, 1175-1180, 2001
A:Title: Cloning and characterization of the cpyA gene encoding intracellular carboxy
A:Reference number: JC7666; MUID:21333188; PMID:11440134
A:Accession: JC7666
A:Molecule type: mRNA
A:Residues: 1-552 <OHS>
A:Cross-references: DDBJ:AB051820
C:Comment: This protein is an intracellular carboxypeptidase, which is localized in v
C:Genetics:
A:Gene: cpyA
A:Introns: 165/1
F:1-15/Domain: signal sequence (or 1-18, or 1-19) #status predicted <SIG>
F:16-131/Domain: pro-sequence, vacuolar targeting signal (or 19-131, or 20-131) #stat
F:131-132/Region: cleavage site, by Arg-Ile #status predicted

Alignment Scores:
Pred. No.: 7.06e-33 Length: 552
Score: 601.00 Matches: 154
Percent Similarity: 47.34% Conservative: 77
Best Local Similarity: 31.56% Mismatches: 175
Query Match: 20.70% Indels: 82
DB: 2 Gaps: 19

US-09-712-338-1_COPY_55_1662 (1-1608) x JC7666 (1-552)
QY      79 GCAACAATGTCCACCATCGGTACAAGGAACCCGGGAGAGGGCGTCTGCCAGACTACC 138
Db      127 AlaTyrAspLeuArgIleLysLysThrAspProSerSerLeuGlyIle-----Asp 143
QY      139 CCGGTGTCAATTCCTACTCTCGATATGTCGACACTCTCCG---GAGTCCCATACCTTC 195
Db      144 ProAspValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 163
QY      196 TTCTGCTCTTCGAGCCAGACATACCCAGAACTGCACCTATCACATTGTGGTTGAT 255
Db      164 TyrTrpPheGluSerArgAsnAspProLysAsnAspProValValLeuTrpLeuAsn 183
QY      256 GSTGCCCTTGGAGGATTTCTTGTGCTCTCTTGAAGAGTTCGGCCCTTGCCTATGTC 315
Db      184 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 203
QY      316 AATTCGACTTTGATGATACATCAACCTCCTGCTGGAACGAGGTCTCCAAATTTACTA 375
Db      204 AspGluAsnIleLysProValTyrAsnProTyrAlaTrpAsnSerAsnAlaSerValIle 223
QY      376 TTCTGTGTCAGCCATTCGGAGTGGCTTTTCATATAGTATACGTTGATGGTTCATT 435
Db      224 PheLeuAspGlnProValAsnValGlyTyrSerTyrSer----- 236
QY      436 AACCCIGTAACCTGGGTGCTCGAAATTCGAGCTTTCAGGAGTTCAGGCGCGGTACCCA 495

```



```

QY 448 GGGTCCTGCAAAATTCAGAGTTTCAGGAGTTTCAGGCGCGGTACCAACCAATTGAGCC 507
Db 231 -----Ser 231
QY 508 ACTCTGATCATCTACCAATCTTGGCGCAGAGCGCGCTTGGGAGATCTCGCAAGGATTC 567
Db 232 GlnSerValSerAsnThrIleAlaAlaGlyLysAspValTyrAlaPheLeuGlnLeuPhe 251
QY 568 CTTAGTGGCACTACCTAGCTCTAGGACTCTAGGCTCAGCTCAGGACITTCAGTCTATGGAGC 627
Db 252 PheLysAsnPro-----GluTyrAlaAsnLeuAspPheHisIleAlaGly 267
QY 628 GAGAGCTATGCGAGGCACTATGCTCGCATTTTCAATCATTTTACGAGCAGAAATGAG 687
Db 268 GluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerGluIleLeuThrHisProGlu 287
QY 688 AGAATTGCCAACGGTAGTGTATAGTGTTCAGCTTAATTTCAACTCTCTGGGAATTAIT 747
Db 288 Arg-----AsnPheAsnLeuThrSerValLeu 296
QY 748 -----AAGGCAATCATCGACGAGCGATCCAGGCGCCCTTACACCTGAAITTCGTGIG 801
Db 297 IleGlyAsnGlyLeuThrAspProLeuValGlnTyrGluTyrTyrGluProMetAlaCys 316
QY 802 ACAAATACCTTACGGTATCAAGGCTGTCAAGCAGACGCTCTACAACTACATCAAGTTTGC 861
Db 317 GlyGluGlyGlyGluProSerValLeuGluProGluGluCysAspGlyMet-----Leu 334
QY 862 AACCAATGCCAAATGTTGCCAGGATTTGATTTCACCTGCAACAGCAGAAACCCGACCC 921
Db 335 AsnLeuLeuProArg---CysLeuSerLeuIleGluSerCysTyrGluSer----- 350
QY 922 GCATTAGCTGACTACGCCCTCTCGCGGAGGACCAACATCTCGAGGGACAAATGTTGAG 981
Db 351 -----GlySerValTrpSerCysValProAlaThrIleTyrCysAsnAsnGlyGlnMet 368
QY 982 GGGCCATACCTAGCGCTTCTGCTGGTGGTGTATGATATATCGGCATCCATCATGATGAC 1041
Db 369 GlyProTyrGlnLys---ThrGlyArgAsnValTyrAspIleArgThrMetCysGluGly 387
QY 1042 CCG-----ACTCGGCCAAGTTATACACAAATTTCTGGCAAGGACTCTGTC 1089
Db 388 SerSerLeuCysTyrSerGlnLeuGluTyrIleAspGlnTyrLeuAsnLeuProGluVal 407
QY 1090 ATGGAGCTATCGCGGTCAACATCAAC---TACACCCAGTCCAAATACAGCTCTACTAC 1145
Db 408 LysLysAlaLeuGlyAlaGluValAspGluTyrGlnSerCysAsnPheAspIleAsnArg 427
QY 1147 GCTTCCAGCAACACGAGGCACTTCTCTGGCCC-----AACTTCATCGAAGACCTC 1197
Db 428 AsnPheMetPheAlaGlyAspTrpMetLysProTyrGlnLysAsnValIleAspLeuLeu 447
QY 1198 GAGGAGATCTTGGCTCTCCCGTGGTGTCTCCCTCATCTAT---GGCGACGCCGATATAC 1254
Db 448 GluLys-----GluLeuProVal-----LeuIleTyrAlaGlyAspLysAspPhe 462
QY 1255 ATCTGCAACTGTTGGCGGTTCAGGCGGTTTCCCTCGTGGCACTACTCCCAAGCGCCG 1314
Db 463 IleCysAsnTrpLeuGlyAsnGlnAlaTrpThrAsnArgLeuGluTrpSerGlySerLys 482
QY 1315 CAGTTCCGAAGCGCGGTATACGCGCCCTGAAAGTCAACGGCGCTCGAGTATGGGAAACT 1374
Db 483 GlyPheThrLysAlaProValLysSerTrpLysValGlyLysAsnAlaAlaGlyGluVal 502
QY 1375 CCGAGATATGTAATTTCTCTACTCGCGTCTATGAGCGGCGCATGAGTCCCATAC 1434
Db 503 LysAsnTyrLysHisPheThrPheLeuArgValPheGlyGlyGlyHisMetValProTyr 522
QY 1435 TACAGCCCATCGCCTCCCTGCAATTTGTTAACCGGACTATCTTCGGI 1482
Db 523 AspGlnProGluAsnAlaLeuAspMetValAsnArgTrpIleSerGly 538

```

RESULT 7

```

S61713
carboxypeptidase C (EC 3.4.16.5) precursor - yeast (Pichia pastoris)
N:Alternate names: carboxypeptidase Y
C:Species: Pichia pastoris
C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 16-Jun-2000
C:Accession: S61713
R:Ohtani, H.; Ohtani, W.; Okazaki, N.; Furuhashi, N.; Ohmura, T.
Yeast 12, 31-40, 1996
A:Title: Cloning and characterization of the Pichia pastoris PRC1 gene encoding carbo
A:Reference number: S61713; MUID:96381245; PMID:8789258
A:Accession: S61713
A:Molecule type: DNA
A:Residues: 1-523 <OH>
A:Cross-references: EMBL:X87987; NID:g1171615; PIDN:CAA61240.1; PID:g1171616
C:Superfamily: serine carboxypeptidase
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-106/Domain: propeptide #status predicted <PRO>
F:107-522/Product: carboxypeptidase C #status predicted <MAT>
F:193,271,484,487/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:249,441,498/Active site: Ser, Asp, His #status predicted
F:296-310,327-336/Disulfide bonds: #status predicted

```

Alignment Scores:

```

Pred. No.: 1.92e-32 Length: 523
Score: 594.50 Matches: 153
Percent Similarity: 45.89% Conservative: 59
Best Local Similarity: 33.12% Mismatches: 169
Query Match: 20.47% Indels: 81
DB: 1 Gaps: 16

```

US-09-712-338-1_COPY_55_1662 (1-1608) x S61713 (1-523)

```

QY 148 AAATCTACTCTCGATATGTCGACAC---TCTCCGAGTCCCATCACTCTCTCTGCTC 204
Db 121 LysGlnTyrSerGlyTyrLeuAspValGluAlaAspAspLysHisPheTyrTrpPhe 140
QY 205 TTCGAGCCAGACATACCCAGAACTGCACCTATCACATTTGGTGTGAATGGTGGCCCT 264
Db 141 PheGluSerArgAsnAspProGlnAsnAspProIleIleLeuTrpLeuAsnGlyGlyPro 160
QY 265 GGAAGCGATTTCTTGATCGTCTCTCGAAGAGTTCGGCCCTGCCATGTCATTCGACT 324
Db 161 GlyCysSerSerLeuThrGlyLeuPhePheGluLeuGlySerSerArgIleAsnGluAsn 180
QY 325 TTGATGACTACATCAACCTCACTCGTGAACGAGGTCTCCAAATTTACTATTCTCTGTC 384
Db 181 LeuLysProIlePheAsnProTyrSerTrpAsnGlyAsnAlaSerIleIleLysLeuAsp 200
QY 385 CAGCCATTTGGAGTCCGCTTTTCATATAGTATACGGTTGATGGGTTCATTACCCCTGTA 444
Db 201 GlnProValAsnValGlyPheSerTyrSer----- 210
QY 445 ACTGGGTCGTCGAAATTCAGCTTTGTCAGGAGTTCAGGCGGTACCAACCACTTGA 504
Db 210 ----- 210
QY 505 GCACTCTGATGATACACTACCAATCTTCCGCGAGAGCGCGCTTGGGAGATCTCGCAAGGA 564
Db 211 SerSerSerValSerAsnThrValValAlaGlyLysAspValTyrAlaPheLeuGlnLeu 230
QY 565 TTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTTCAGTCTAAGACATCTAGTCTAGTGG 624
Db 231 PhePheGlnHisPhePro-----GluTyrGlnThrAsnAspPheHisIleAla 246
QY 625 ACGGAGAGCTATGAGGCGACCATGCTCTGCTGCAATTCCTCAATCATTTTACAGCAGAA 684
Db 247 GlyGluSerTyrAlaGlyHisTyrIleProValPheAlaAspGluIleLeuSerGlnLys 266
QY 685 GAGAGAATTGCCAACGGTAGTGTATTAATGTTGCTCAGCTTAATTTCAACTCTCTGGGAAT 744
Db 267 AsnArg-----AsnPheAsnLeuThrSerVal 275

```


296	AsnArgTyr	14	Gaps:	14
901	TGCAACAGACAAACCCAGCGCATTAAGTACTGACTACGGC	1662	(1-1608) x A35275 (1-411)	
302	CysGlnGlnSerIleLeuGluCysSerAlaAspGlyGlyGluGlyAspAlaCysAlaSer	165	CCCCGGG--GCAGAGGGGTCTGCGAGACTACCCCGGGTCAAAATCTACTCTGATAT	
952	GCACCAACATCTGACGAGCAATGTTGAGGGCCATACACTAGCGCTTCTGCTGCTGCT	17	ProGlyLeuProGluGlyValAlaAsp-----LeuGlyHisHisAlaGlyTyr	
332	SerTyrThrValCys--AsnAsnIlePheGlnIleMetAspIleAlaGlyAsnVal	219	GTGCACACCTCTCCGAGTCCCATACCTCTTCTGGTCTTCTGAGGACGACAT	
1009	GGTGTGTATATTCGGCTCCATATGATGACCCGACTCCGCCAAGTTATTACAAAC---	37	TyrArgLeuProAsnThrHisAspAlaArgMetPheTyrPhePheGluSerArgGly	
341	AsnTyrTyrAspValArgLysGlnCysGluGlySerLeuCysTyrAspPheSerAsnMet	279	ACCCGAGAACTGACCATATCATCTGTTCAATGCTGGCCTCGAAGCGATCTTTG	
1066	---AAATTTCTGGCAAGACTCTGTCATGGACGCTATCGGGCTC---AACATCACTAC	56	LysLysGluAsp---ProValValIleTrpLeuThrGlyGlyProGlyCysSerSerGlu	
361	GluAsnPheLeuAsnGlnLysSerValArgLysAlaLeuGlyValGlyAspIleGluPhe	339	ATCGGCTCTTCGACAGATGGCCCTGTCATGTCATTAATTCGACATTTGATGACTATC	
1120	ACCCAGTCCAAATAGCTCTACTACGCTTTCCAGCAACAGCGGCTTTGCTGGCC	76	LeuAlaValPheTyrGluAsnGlyProPheThrIleAlaAsnAsnMetSerLeuValTrp	
381	ValSerCysSerThrAlaValIleGluAlaMetGlnMetAsp-----TrpMet	399	AACCTCTACTCGTGGACGAGTCTCCAAATTTACTATTCTCTGCCAGCATTTGGGAGTC	
1180	AACTTCATGAA---GACCTCGAGGAGATCTCTCCCGTGGCTCTCCCTCATC	96	AsnLysPheGlyTrpAspLysIleSerAsnIlePheValAspGlnProThrGlyThr	
397	ArgAsnLeuGluValGlyIleProAlaLeuLeuGlnAspGlyIleLysLeuLeuValTyr	459	GGCTTTTCATATAGTATAGTGGTGAUGGGTCCATTAAACCTGTAACCTGGGCTCGTGA	
1237	TATGGCGGCGGATTCATCTCACTCGCTGTTGCGGGTTCAGGCGCTTTCCCTCGCTCG	114	GlyPheSerTyrSerSerAspAspArgAspThrArgHisAspGluThrGlyVal----	
417	AlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerLysTrpValHisGluMet	519	AATTCGAGCTTTCAGGAGTTCAGGGCGGTACCAACCAATGATGACCATCTGATCGAT	
1297	AACCTCTCCCAAGCCCGCTGTCGAGAGCGGTACAGCGCCCTGAAAGTCAACGGC	114	-----	
437	GluTrpSerGlyGlnLysGluPheValAlaAlaAlaThrValProPheHisValAspAsn	579	ACTACCAATCTTCCGAGAGCGCGTGGAGATCTGCAAGATTCCTTAGTGGACTA	
1357	GTCAGTATGGGAACTCGAGTATGTTATTTCTCTACTCGCTCTATGAGGCA	129	-----SerAsnAspLeuTyrAspPheLeuGlnValPhePheLysLysHis	
457	LysGluAlaGlyLeuMetLysAsnTyrGlySerLeuThrPheLeuLysValHisAspAla	639	CCFAGCTTGGACTCTAGGTCACCTCTAAGGACTCAGTCTATGAGCGAGACTATGGA	
1417	GGCCTAGTCCCATCTACAGCCCATCGCTCTCCCTCAATGTTTACCGGACTATC	145	ProGluPhe-----IleLysAsnAspPhePheIleThrGlyGluSerIleAla	
477	GlyHisMetValProMetAspGlnProLysAlaAlaLeuGlnMetLeuGlnAsnTrpMet	699	GGGCACTATGCTCTCATCAATCTTACGAGCAAGATGAGAAATGCGCAAC	
1477	TTCGGT 1482	162	GlyHisIleProAlaPheAlaSerArgValHisGlnGlyAsnLysLys-----	
497	GluGly 498	759	GGTAGTCTTAATGCTGCTAGCTTAATTTCAACTCTCTGGAATATTAACGGCATCATC	
		180	-----AsnGluGlyThrHisIleAsnLysGlyPheAlaIleGlyAsnGlyLeuThr	
		819	GACGAGGATCCAGGCGCTTACTACCTGAATTCGCTGTGAACAATACCTACGGTATC	
		200	AspProAlaIle***TyrLysAlaTyrThrAspTyrAlaLeuGluMetAsnLeuIleGln	
		879	AAGCTGTCAACGAGACGCTCTACAACTACATGAAGTTTGCACCAACCAATGCCAAATGGT	
		213	LysAlaAspTyrGluArgIle-----AsnLysPheIleProPro	
		936	TGCCAGGATTTGATTTCCACTGCAACAGACAAACCCG---ACCGAATTTAGCTACAC	
		233	CysGluPheAlaIleLysLeuCysGlyThrAsnGlyLysAlaSerCysMetAlaAlaTyr	
		996	GCCTCTCGCGGAGCCCAACATGTGCGAGGACAATGTGAGGGGCCCTACTACGCC	
		252	MetValCysAsnThrIlePheAsnSerIleMetLysLeuValGlyThrLysAsnTyr---	
		1056	TTTGTGCTGCTGCTGCTATGATATTCGGATCCATATGATGAGCCCACTCCGCAAGT	
		266	-----TyrAspValArgLysGluCysGlyLysLeuCysTyrAsp	
		1107	TATTACAAC-----AAATTTCTGCAAGGACTCTGTCATGAGCGTCTACCGGTC---	
		286	PheSerAsnLeuGluLysPhePheGlyAspLysAlaValArgGluAlaIleGlyValGly	
		1167	RACATCACTACACCCAGTCCCAATATGACGCTCTACTACGCTTTCCAGCAACAGCGGAC	

RESULT 12

A35275
carboxypeptidase C (BC 3.4.16.5) - barley
N: Alternate names: carboxypeptidase III
C: Species: Hordeum vulgare (barley)
C; Date: 24-Nov-1999 #sequence, revision 24-Nov-1999 #text_change 24-Nov-1999
R: Sorensen, S.B.; Svendsen, I.; Breddam, K.
A: Title: Primary structure of carboxypeptidase III from malted barley.
A; Reference number: A35275; MUID: 90315015; PMID: 2639682
A; Accession: A35275
A; Molecule type: protein
A; Residues: 1-411 <SOR>
A; Note: 185-Val was also found
C; Superfamily: serine carboxypeptidase
C; Keywords: blocked amino end; glycoprotein; hydrolase; serine carboxypeptidase
F; 1/ Modified site: blocked amino end (Leu) (probably acetylated) #status experimental
F; 7/ Binding site: carbonyl group (Asn) (covalent) #status experimental
F; 143, 331, 388/ Active site: Ser, Asp, His #status predicted

Alignment Scores:			
Pred. No.:	5, 1e-21	Length:	411
Score:	425.50	Matches:	124
Percent Similarity:	42.21%	Conservative:	71
Best Local Similarity:	26.84%	Mismatches:	200
Query Match:	14.65%	Indels:	67

```

Db      287  AsnLeuGluPheValSerCysSerThrSerValTyrGlnAlaMetLeuThrAsp----- 304
QY      1168  TTTCCTGCGCCCAACTTCATCGAA--GACCTCGAGGAGATCTTCTCTCCCGGTGGT 1224
Db      305  -----TrpMetArgAsnLeuGluValGlyLeuProAlaLeuLeuGluAspGlyIleAsn 322
QY      1225  GTCCTCCCTCACTATGCGGAGCGGATACATCTGCACTGCTCGGCGGTGAGCCGTT 1284
Db      323  ValLeuIleTyrAlaGlyLeuTyrAspLeuIleCysAsnTrpLeuGlyAsnSerArgTrp 342
QY      1285  TCOCTCGCTGGCACTACTCCAGCCCGCCAGTTCGGAAGCGAGGTACAGCCCGCTG 1344
Db      343  ValHisSerMetGluTrpSerGlyGlnLysAspPheAlaLysThrAlaGluSerSerPhe 362
QY      1345  AAAGTCAACGGCGTCTGATGATGGGAACCTCGCGAGTATGTAATTCCTCTCACTCGC 1404
Db      363  LeuValAspAspAlaGlnAlaGlyValLeuLysSerHisGlyAlaLeuSerPheLys 382
QY      1405  GTCATGAGGAGCGCCATGAAGTCCCATCTACTACCGCCATCGCTCCCTGCAATGTTT 1464
Db      383  ValHisAsnAlaGlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeu 402
QY      1465  AACCGG 1470
Db      403  ArgArg 404

```

RESULT 13

A29412

carboxypeptidase C (EC 3.4.16.5) precursor - wheat

N:Alternate names: carboxypeptidase Y homolog

C:Species: Triticum aestivum (common wheat)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Nov-1999

C:Accession: A29412

R:Baulcombe, D.C.; Barker, R.F.; Jarvis, M.G.

J. Biol. Chem. 262, 13726-13735, 1987

A:Title: A gibberellin responsive wheat gene has homology to yeast carboxypeptidase Y.

A:Reference number: A29412; MUID:88007602; PMID:2820978

A:Accession: A29412

A:Molecule type: mRNA

A:Residues: 1-499 <BAU>

C:Genetics:

A:Introns: 100/2; 156/3; 244/3; 295/1; 324/3; 411/1; 455/3

C:Superfamily: serine carboxypeptidase

C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase; zymogen

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-73/Domain: propeptide #status predicted <PRO>

F:74-484/Product: carboxypeptidase C #status predicted <MAT>

F:485-499/Domain: carboxyl-terminal propeptide #status predicted <CPRO>

F:144/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:216,404,461/Active site: Ser, Asp, His #status predicted

Alignment Scores:

```

Pred. No.:      3,47e-20      Length:      499
Score:          413.50      Matches:     126
Percent Similarity: 41.37%      Conservative: 80
Best Local Similarity: 25.30%      Mismatches: 217
Query Match:      14.24%      Indels:      75
DB:              1          Gaps:       16

```

US-09-712-338-1_COPY_55_1662 (1-1608) x A29412 (1-499)

```

QY      1  CTTCCAGGAGTACACCGCGCTCGGTGAGTACAGACAGCTACCCCAAGAACCCCGG 60
Db      47  LeuProGlyArgProArgArgGlyLeuGly-----AlaGly 58
QY      61  GTCAGACTCTTACACCCCAACAAATGTCACATCCCGTACAGGAACCCGG---GCA 117
Db      59  AlaGluAspValAlaProGlyGlnLeuLeuGluAlaArgValThrLeuProGlyLeuPro 78
QY      118  GAGGCGTCTCGAGACTACCCCGGTGTCAATCTCTACTCTGGATAT-----GTGAC 171
Db      79  GluGlyValGlyAsp-----LeuGlyHisHisAlaGlyTyrTyrArgLeuPro 94

```

```

QY      172  ACCCTCCCGAGTCCCATACCTCTCTCTGTTCTTCCGAAGCCAGACATAACCCAGAACT 231
Db      95  AsnThrHisAspAlaArgMetPheTyrPhePheGluSerArgGlyLysGluAsp 114
QY      232  GCACCTATCAATGTGTTGAATGTCGCCCTCGAAGCGATTTTGTTCGTCCTCTTC 291
Db      115  ---ProValValIleTrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaValPhe 133
QY      292  GAAGAGTTGGCGCTTGCCATGTCATCTTGTATGATGATCACTCAACCCCTCACTCG 351
Db      134  TyrGluAsnGlyProPheThrIleAlaAsnMetSerLeuValTrpAsnLysPheGly 153
QY      352  TGGACAGAGTCTCCAAATTTACTATTCTGTCGCCAGCCATTCGGAGTGGCTTTTCAT 411
Db      154  TrpAspLysIleSerAsnIlePheValAspProAlaThrGlyThrGlyPheSerTyr 173
QY      412  AGTGATACGGTGTGATGGTCCATTAACCCCTGTAAGTGGGTGGTGGAAATTCGAGCTT 471
Db      174  SerSerAspAspArgAspThrArgHisAspGluAlaGlyVal----- 187
QY      472  GCAGGAGTTCAGGGCGGTACCCCAACCATTCGACCACTCTGATCGATACCTACCACT 531
Db      187  ----- 187
QY      532  GCCGACAGAGCGCGCTTGGGAGATCTGCAAGATTCCTTAGTGGACTACCTAGCTTGG 591
Db      188  ---SerAsnAspLeuTyrAspPheLeuGlnValPhePheLysHisProGluPhe--- 205
QY      592  ICTAGGGTGCACCTTAAGGACTTCAGTCTATGAGGAGAGCTATGGAGGCGACTATGGT 651
Db      206  -----ValLysAsnAspPheIleThrGlyGluSerLysAlaGlyHisTyrIle 222
QY      652  CCTGCATCTTCAATCTTTACGACGACGAATGAGAGAAATGCCACGGTAGTGTATAT 711
Db      223  ProAlaPheAlaSerArgValHisGlnGlyAsnLysLys-----AsnGlu 237
QY      712  GGTCTTCAGCTTAATTTCAACTCTCTGGGATTTATACGGCATCATCGAGGCGCATC 771
Db      238  GlyThrHisIleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaIle 257
QY      772  CAGSCCCTTACTACCCGATTCGCTGTGAACAATACCTACGATATCAAGGTGCTCAAC 831
Db      258  GlnTyrLysAlaTyrThrAspTyrAlaLeuAsp-----MetAsnLeuIleGln 273
QY      832  GAGACCGCTCTACAACTACATGAAGTTTCCCAACCAATGCCAAATGTTGCCAGGATTG 891
Db      274  LysAlaAspTyrAspArgIle-----AsnLysPheIleProProCysGluPheAla 290
QY      892  ATTTCCACCTGCAACAGACAAACCCGC---ACCGCATTAGCTACGCGCTCTGCGCC 948
Db      291  IleLysLeuCysGlyThrAspGlyLysAlaSerCysMetAlaAlaTyrMetValCysAsn 310
QY      949  GAAGCCACCAACATGTCAGGGACAATGTTGAGGGCCATACCTACGCGCTTGTGTCGT 1008
Db      311  SerIlePheAsnSerIleMetLysLeuValGlyThrLysAsnTyr----- 325
QY      1009  GGTGTGTATGATATTCGGCATTCATATGATGACCGCATCTCCGCCCAAGTATTACAAC 1065
Db      326  -----TyrAspValArgLysGluCysGluGlyLysLeuCysTyrAspSerAsnLeu 343
QY      1065  ---AAATTCTGCCAAGGACTCTGTCTATGGAGCTATCGCGTC---AACATCAACTAC 1119
Db      344  GluLysPhePheGlyAspLysAlaValArgGlnAlaIleGlyValGlyAspIleGluPhe 363
QY      1120  ACCCAGTCCAAATGACGCTACTACGCTTCCAGCAACAGCGGACTTGTCTGGCCC 1179
Db      364  ValSerCysSerThrSerValTyrGlnAlaMetLeuThrAsp-----TrpMet 379
QY      1180  AACTTCATCGAA---GACCTCGAGGAGATCTCTGCTCTCCCGTGGTCTCTCCCTCATC 1236
Db      380  ArgAsnLeuGluValGlyIleProAlaLeuLeuGluAspGlyIleAsnValLeuTyr 399

```

1237 TATGCGCAGCCGGATTACACTCTCAACTGGTTCGGCGGTACGCCGGTTTCCTCGCTGGC 1296
 QY ||||| ||| ||||||||| ||| :
 Db 400 AlaGlyGluTyrAspLeuileCysAsnTrpLeuglyAsnSerArgTrpValHisSerMet 419
 QY ||||| ||||| ||||| ||||| ||||| :
 Db 1297 RACTACTCCCAAGCGCCCGACGTTCGGAAGCGCAGGTACACGCCCTCGAAAGTCAACGGC 1356
 QY ||||| ||||| ||||| ||||| ||||| :
 Db 420 GluTrpSerGlyGlnLysAspPheAlaLysThrAlaGluSerPheLeuValAspAsp 439
 QY ||||| ||||| ||||| ||||| ||||| :
 Db 1357 GTCGAGTATGGGAACACGCGGAGPATGGTAATTTCTCTTCACTCGCGTCTATCAGGCA 1416
 QY ||||| ||||| ||||| ||||| ||||| :
 Db 440 AlaGlnAlaGlyValLeuLysSerHisGlyAlaLeuSerPheLeuLysValHisAsnAla 459
 QY ||||| ||||| ||||| ||||| ||||| :
 Db 1417 GCCCATGAAGTCCCATAFACACGCCCATCGCTCCCTCCGCAATGTTTAAACCGG 1470
 QY ||||| ||||| ||||| ||||| ||||| :
 Db 460 GlyHisMetValProMetAspGlnProIysAlaAlaLeuGluMetLeuArgArg 477

RESULT 14

T18968
 probable serine-type carboxypeptidase (EC 3.4.16.-) Y16B4A.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Mar-2000
 C:Accession: T18968; T23145; T26477
 R:Thomas, K.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: Z19053
 A:Accession: T18968
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-2105 <W1>
 A:Cross-references: EMBL:Z70203; PIDN:CAA94110.1; GSPDB:GN000028; CESP:Y16B4A.2
 A:Experimental source: clone C05G5
 R:Lloyd, C.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19697
 A:Accession: T23145
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-2105 <W1>
 A:Cross-references: EMBL:Z93339; PIDN:CAB07544.1; GSPDB:GN000028; CESP:Y16B4A.2
 A:Experimental source: clone H40D05
 R:Wallis, J.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: Z20220
 A:Accession: T26477
 A>Status: preliminary; translated from GB/EMBL/DDEJ
 A:Molecule type: DNA
 A:Residues: 1-2105 <W1>
 A:Cross-references: EMBL:AL023825; PIDN:CAA19443.1; GSPDB:GN000028; CESP:Y16B4A.2
 A:Experimental source: clone Y16B4A
 C:Comment: This protein has 4 repeats of sequence homologous to serine carboxypeptidase.
 C:Genetics:
 A:Gene: CESP:Y16B4A.2
 A:Map position: X
 A:Introns: 43/3; 67/2; 179/1; 208/3; 239/2; 281/2; 296/3; 346/2; 444/3; 503/3; 588/3; 68
 1664/3; 1752/2; 1794/3; 1834/3; 1882/2; 1905/2; 1942/3; 1982/3
 C:Keywords: duplication; hydrolase; serine carboxypeptidase

Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
4, 79e-20	2105	413.50	Matched: 150
Conservative:	85		
Percent Similarity:	38.52%		
Best Local Similarity:	24.59%		
Query Match:	14.24%		
Indels:	173		
Gaps:	26		

US-09-712-338-1_COPY_55_1662 (1-1608) x T18968 (1-2105)

QY 10 AGTACACCGGTCGTCGGTAGACAGACGTACCACAGAACCCCCCGGGTCAAGACT 59
 ||||| ||||| ||||| :
 Db 1521 SerThrProAlaGlyle-----AspValThrProglinProLeuThrAsn 1535
 QY 70 CTTACAACCGCAACAANTCTACCATCGGTACAGGAACCCCGGGGCAGAGGGCTCTGC 129

```

QY 1069 -----TTTCIG 1074
Db 1841 SerSerAspPheLeuAsnGlyTyrProCysTyrAspSerAlaMetGluGlyTyrLeu 1860
QY 1075 GCAAGGACTCTGTCATGGCGCTATCGCGTC-----AACATCAACIACACCCAGTCC 1128
Db 1861 AsnArgProValValArgLysAlaLeuAsnIleProAspSerValProTyrTyrAlaAla 1880
QY 1129 AATAATGACGTCTACTACGCTTTCCAGCAACAGCGCCTTTGTCIGCGCCCACTTCAIC 1188
Db 1881 AsnAsnIleIleAsnAlaTyrAsnGlnValAspSerIleValProAsn----- 1898
QY 1189 GAAGACCTCGAGAGATCCTTGT-----CTCCCGTGGCTGCTCCCTCAICAT 1239
Db 1899 -----LeuGlnIleIleMetAlaAsnAlaProAlaAsnPheLysMetLeuLeuTyrSer 1916
QY 1240 GCGGAGCGCGATTCATCTCAACTGCTTCCAGGCGGTTCAG----- 1278
Db 1917 GlyAspAlaAspThrMetValAsnTrpLeuGlyAlaGluIlePheThrAlaAsnAsnPhe 1936
QY 1279 ---GCCGTTTCCTCGCTGCGAACTATCTCCAGCGCGCC-----CAGTTCGGAAGC 1326
Db 1937 AlaAlaLeuGlyLeuThrThrSerSerProArgAlaGlnTrpThrPheGlnTyrAsnSer 1956
QY 1327 GCAGGTACACGCGCTGAAAGTCACGCGCTCGAGTATGGGAACTCGCGAGTATGCT 1386
Db 1957 Thr---PheGlnPro---ThrValAlaGlyTyrGlnThrSerTyrThrSerAsnAlaIle 1974
QY 1387 AAPTTCCTCTCACTCGCTCTATGAGGAGCGCATGAGTCCCATCTACTACGAGCCCATC 1446
Db 1975 AsnIleAspValLeuThrValLysGlySerGlyHisPheValProLeuAspArgProGln 1994
QY 1447 GCCTCCTCGAATGTTTAAACCGGACTATCTTCGGTGGGATATCGCAGAGGCCAGAG 1506
Db 1995 GlnAlaLeuGlnMet-----IleTyrAsnPheValLysSerArgGlyTyrAsn 2010
QY 1507 AAGATCTGCGCCGACGACAGAGATGGAAGCGGTACAGCTACGACATACAGCTGCTCC 1566
Db 2011 ThrProPheAspLeuAsnSerAsnPheThrThrThrThrThrThrThrThrThrThrThr 2030
QY 1567 GTGCG-----CTGCGTACGCTACACG 1590
Db 2031 ThrProGlyThrGlyProThrValThrAla 2040

RESULT 15
serine-type carboxypeptidase like protein - Arabidopsis thaliana
N:Alternate names: protein F4F15.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T49079
R:Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25015
A:Accession: T49079
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <ALC>
A:Cross-references: EMBL:AL049711; GSPDB:GN00061; ATSP:F4F15.110
A:Genetic:
A:Experimental source: cultivar Columbia; BAC clone F4F15
A:Gene: ATSP:F4F15.110
A:Map position: 3
A:Introns: 116/1; 151/1
C:Superfamily: serine carboxypeptidase

Alignment Scores:
Pred. No.: 4e-18 Length: 482
Score: 383.00 Matches: 135
Percent Similarity: 39.42% Conservative: 68
Best Local Similarity: 26.21% Mismatches: 202
Query Match: 13.19% Indels: 110

```

```

DB: 2 Gaps: 18
US-09-712-338-1_copy_55_1662 (1-1608) x T49079 (1-482)
QY 43 CCCAAGAACCCACCGGTCAGACTCTTACACCGCAACAATGTCCACATCCGGTAC 102
Db 32 ProSerProLysArgGlyValSerSerSerGlyAspThrSerHisPheAsnValIleGln 51
QY 103 AAGGAA-----CCCGGGCGAGAGCGGCTCGCGAGACTACCCCGGCT----- 144
Db 52 ArgGluSerValProSerProLysAspLysLeuIleGlnLeuProGlyGlnPro 71
QY 145 -----GTCAATCTCTACTGGAATATGCGACACTCTCCGAGTCCCATACC 192
Db 72 SerAspValThrPheLysGlnIyrGlyGlyTyrValAlaValAsnLysProAlaGlyArg 91
QY 193 TTC-----TTCCTGCTTCTCGAAGCCAGACATACCCAGAACTGCACCTATCACATTG 246
Db 92 PheLeuTyrTyrIyrPheValGluThrIleLysProGlyAsnThrThrProLeuValIle 111
QY 247 TGGTGTGAAGTGGCCCTGGAAGCATTCCTTTGATCGGCTCTCTTCGAAGAGTTCGGCCCI 306
Db 112 TrpPheAsnGlyGlyProGlyCysSerSerLeuGlyGlyAlaPheLysGlnLeuGlyPro 131
QY 307 TGCATGTCAATTCGACTTTTGATGAC---TACATCAACCTCCTACTCGTGGAAACGAGGTC 363
Db 132 PheArgValHisSerAspGlyLysThrLeuPheArgAsnProTyrSerTrpAsnAsnGlu 151
QY 364 TCCAATTTACTATTCCTGTCGCCAGCCATTCGGAGTGGCTTTTCATATAGTATGATACGGTT 423
Db 152 AlaAsnValLeuPheLeuGluThrProValGlyThrGlyPheSerTyrSer----- 168
QY 424 GATGGTCCATTAACCTGTAACTGGGCTGCTCGAAAATTCGAGACTTCGAGGAGTTCAG 483
Db 169 -----AsnSerProIleAsnGlyLysGln 176
QY 484 GGCCTGTACCCACCATTCATGATGATGCTCTAGTGGACTACCTAGCTTGGACTAGGGTGCAG 543
Db 177 GlyAspLysAlaThr-----AlaGluAsp 184
QY 544 GCTGGGAGATCCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTAGGGTGCAG 603
Db 185 AsnTyrMetPheLeuValAsnTrpLeuGluArgPheProGluTyrLysGlyArg----- 202
QY 604 TCTAAGGACTTCAGTCTATGAGCGAGAGCTATGAGGCGCAGTATGCTGCTCATCTTC 663
Db 203 -----AspIleTyrIleAlaGlyGlnSerTyrAlaGlyHisTyrValProGlnLeuAla 220
QY 664 AATCATTTTACGAGCAGATGAGAGAAATGCCACGGTAGTGTAAAGTGTTCAGCTT 723
Db 221 GlnIleIleLeuHisArgAsn-----AsnGlnThrLeuIle 232
QY 724 AATTTCACATCTCTGGGAATTATTACGGCATCATCGAGCGGATCCAGGCCCTTAC 783
Db 233 AsnLeuArgGlyIleLeuIleGlyAsnProSerLeuAsnArgGluIleIleAspPhe 252
QY 784 TACCCTGATTCGCTGTGAACAAATACCTACGGTATCAAGGCTGTCAACGAGACCGCTAC 843
Db 253 GlyTyrLysPheMetPhe-----SerHisGlyLeuIleSer---GlnGlnGlnMetAsp 269
QY 844 AACTACATGAAGTTTGCCCAACCAATGCCAAATGTTGCCAGGATTTGATTTCCACCTGC 903
Db 270 AsnTyrAsnLysPhe-----CysThrAsp-----SerAspLeu 280
QY 904 AACACAGACAAACCGCACCGCATACCTGACTACGCCCTCTCGCCGAGACCCACCAATG 963
Db 281 TyrAspTrpAspLysCysHisLeuAlaSerGlnLysIleGluAlaGlnLysThrHisLeu 300
QY 964 TGCAGGGACAATGTTAGGGGCCACTACTACGCCCTTGTGCTGCTGGTGTGTATGATATT 1023
Db 301 -----AspIleIyrAsnIle 305
QY 1024 CGGCATCCATAT----- 1035

```

wed Sep 17 15:23:15 2003

```

Db      306 TyrAlaProLeuCysLeuAsnSerThrLeuSerSerGluProLysLysCysThrThrIle 325
      |||
QY      1036 -----GATGACCGACTCCGCCAAGTATTACAAATAATTTCTGGCAAGGACTCTGTC 1089
      ||||| :||| :||| :||| :|||
Db      326 MetLysAlaAspProCysSerGlyAsnTyrLeuLysAlaTyrLeuAsnIleLysGluVal 345
      ||||| :||| :||| :||| :|||
QY      1090 ATGGACGGTATCGCGGTCAAC-----ATCAACTACACCCAGTCCCAATAAT 1134
      :||| :||| :||| :||| :|||
Db      346 GlnGluAlaIleHisAlaAsnThrThrLysIleProLysGluTyrThrSerCysAsnThr 365
      :||| :||| :||| :||| :|||
QY      1135 GACGTCCTACTACGGCTTTCAGCAACAGCGACTTTGCTGGCCCAACTTCATCGAAGAC 1194
      :||| :||| :||| :||| :|||
Db      366 LysLeuLeuTyrGluTyrPasnGluLysAspArgTyrVal-----SerLeuThrProIle 383
      :||| :||| :||| :||| :|||
QY      1195 CTCGAGGAGATCCTGCTCTCCCGTGGTGTCTCCCTCATCTATGGCGACGCCGATAC 1254
      ||||| :||| :||| :||| :|||
Db      384 LeuGlnGluLeuMetGlyLysGlyValArgValMetLeuTyrAsnGlyAspValAspLeu 403
      :||| :||| :||| :||| :|||
QY      1255 ATCTGCAACTGGTTCGGCGGTACAGCGCGTTTCCCTCGCTGGGAAGTACTCCCAAGCGGC 1314
      :||| :||| :||| :||| :|||
Db      404 ValIleProPheThrSerThrLeuAlaValValLysThrMetAsnLeuThrValValLys 423
      :||| :||| :||| :||| :|||
QY      1315 CAGTTCCGAAGCGAGGTACACGCCCTGAAAGTCAACGGCTCGAGTATGGGAAACT 1374
      :||| :||| :||| :||| :|||
Db      424 GluTyrArg-----ProTyrPheThrGlyGlyHisValGlyGlyPheThr 438
      :||| :||| :||| :||| :|||
QY      1375 CGCGAGTAT---GGTAATTCTCCTTCATCTCGCTCATGAGCGAGCGCATGAAGTCCCA 1431
      :||| :||| :||| :||| :|||
Db      439 GluAspTyrLysGlyAsnLeuThrPheValThrValLysGlyAlaGlyHisSerValPro 458
      :||| :||| :||| :||| :|||
QY      1432 TACTACCGCCATCGCTCCCTGCAATTGTTTAAACCGGACTATC 1476
      :||| :||| :||| :||| :|||
Db      459 ThrAspGlnProIleHisAlaLeuAsnIlePheThrSerPheIle 473
      :||| :||| :||| :||| :|||

```

Search completed: September 16, 2003, 18:48:40
Job time : 71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:04:58 ; Search time 29 Seconds
(without alignments)
5215.099 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 2904
Sequence: 1 cttccaggaagcacccgc.....gcattgcacgtgtgtgtatg 1608

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q-/cg2_1/USPTO_spool_p/US09712338/runat_16092003_144325_14345/app_query.fasta_1.1799
-DB=SwissProt 41 -PMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 -CGN_1_1_24 -runat_16092003_144325_14345 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	30.8	423	1	CPSL_PENJA
2	652.5	22.5	1002	1	CPFY_SCHPO
3	636	21.9	508	1	YBY9_YEAST
4	599	20.6	542	1	CPFY_CANAL
5	594.5	20.5	523	1	CPFY_PICPA
6	527.5	18.2	532	1	CPFY_YEAST
7	474	16.3	516	1	CPFX_ARATH
8	473	16.3	429	1	CPFX_ORYSA
9	452	15.6	500	1	CPFX_ORYSA
10	434	14.9	508	1	CPFX_ORYSA
11	413.5	14.2	500	1	CPFX_ORYSA
12	359	12.4	476	1	CPFX_ORYSA
13	353.5	12.2	436	1	CPFX_ORYSA
14	352.5	12.1	470	1	CPFX_ORYSA
15	351	12.1	482	1	CPFX_ORYSA
16	349	12.0	474	1	CPFX_ORYSA
17	344.5	11.9	516	1	CPFX_ORYSA
18	332	11.4	469	1	CPFX_ORYSA

19	329.5	11.3	480	1	PRTP_HUMAN
20	326.5	11.2	454	1	YUA6_CAEEL
21	321.5	11.1	729	1	KEX1_YEAST
22	318.5	11.0	476	1	CBP2_HORVU
23	316	10.9	574	1	YBP3_CAEEL
24	312.5	10.8	471	1	VCP_AEADA
25	312	10.7	510	1	CBP1_ORYSA
26	305.5	10.5	423	1	CBP2_WHEAT
27	297.5	10.2	505	1	YXD2_CAEEL
28	295.5	10.2	286	1	CBPX_PEA
29	280	9.6	499	1	CBP1_HORVU
30	269.5	9.3	523	1	PEPS_ASPSA
31	255	8.8	452	1	RISC_MOUSE
32	248	8.5	452	1	RISC_RAT
33	237	8.2	452	1	RISC_HUMAN
34	225.5	7.8	507	1	SWA2_SCHPO
35	221.5	7.6	531	1	PEPF_ASPNG
36	206	7.1	5703	1	MUSB_HUMAN
37	195.5	6.7	5179	1	MUC2_HUMAN
38	191.5	6.6	1161	1	DAN4_YEAST
39	190	6.5	1367	1	AMYH_YEAST
40	167.5	5.8	324	1	CP21_HORVU
41	166.5	5.7	1233	1	MUSB_HUMAN
42	166	5.7	907	1	VGP3_EBV
43	160	5.5	725	1	AGAL_YEAST
44	159	5.5	610	1	MUC4_HUMAN
45	157	5.4	797	1	VGLX_RSVEB

ALIGNMENTS

RESULT 1	CPSL_PENJA	STANDARD;	PRT;	423 AA.
ID	CPSL_PENJA			
AC	P34946;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DI	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carboxypeptidase SI (EC 3.4.16.6).			
OS	Penicillium janthinellum (Penicillium vitale).			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;			
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.			
OX	NCBI_TaxID=5079;			
RP	[1]			
RN	SEQUENCE.			
RX	MEDLINE=94039747; PubMed=8224168;			
RA	Svendsen I., Hofmann T., Endrizzi J., Remington S.J., Breddam K.;			
RT	"The primary structure of carboxypeptidase SI from Penicillium			
RT	janthinellum".			
RL	FEBS Lett. 333:39-43(1993).			
CC	-!- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine			
CC	or lysine residue.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.			
DR	PIR; S38953; S38953.			
DR	HSSP; P00729; IYSC.			
DR	MEROPS; S10.008; -.			
DR	InterPro; IPR000379; Ser_estrs_site.			
DR	InterPro; IPR001563; Serine_carbpept.			
DR	Pfam; PF00450; serine_carbpept; 1.			
DR	PRINTS; P00724; CBXOXYPIASEC.			
DR	ProDom; P0001189; Serine_carbpept; 1.			
DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.			
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.			
KW	Hydrolase; Carboxypeptidase; Glycoprotein.			
FT	ACT_SITE 143 143			
FT	ACT_SITE 340 340			
FT	ACT_SITE 397 397			
FT	BINDING 343 343			
FT	BINDING 398 398			
FT	DISULFID 8 68			
FT	DISULFID 55 300			
FT	DISULFID 223 246			
FT	DISULFID 230 239			

P10619 homo sapien
P52715 caenorhabdi
P09620 saccharomyc
P08818 hordeum vul
P52716 caenorhabdi
P42660 aedes aegypt
P37890 oryza sativ
P08819 triticum ae
P52714 caenorhabdi
Q41005 pisum sativ
P07519 hordeum vul
P52719 aspergillus
Q920A5 mus musculu
Q920A6 rattus norv
Q9Hb40 homo sapien
P32825 schizosacch
P52718 aspergillus
Q09hc84 homo sapien
Q02817 homo sapien
P47179 saccharomyc
P08640 saccharomyc
P55747 hordeum vul
P98088 homo sapien
P03200 epstein-bar
P32323 saccharomyc
Q99102 homo sapien
P28968 equine herp

FT	CARBOHYD	200	200	N-LINKED (GLCNAc. . .)
SC	SEQUENCE	423 AA:	45496 MW:	55F33333608B12729 CRC64;

Alignment Scores:	1.02e-54	Length:	423
Pred. No.:	894.00	Matches:	186
Score:	54.31%	Conservative:	66
Percent Similarity:	40.09%	Mismatches:	160
Best Local Similarity:	30.79%	Indels:	52
Query Match:	1	Gaps:	8
DB:			

US-09-712-338-1 COPY 55 1662 (1-1608) x CPS1_PENJA (1-423)

QY	121	GGCGTCTGGAGACTACCCGGGTGTCAAACTCTACTCTGGATATGTCGACACCTCTCCC	180
DB	6	GLYIleCysGluThrThrProGlyValAsnGluTyrSerGlyTyrLeuSerValGlySer	25
QY	181	GAGTCCCATACCTCTCTCTGGTCTTCGGAAGCCAGACATAACCCAGAACTGCACCTATC	240
DB	26	AsnMetAsnMetIrpPheIrpPhePheGluAlaArgAsnAsnProGlnGlnAlaProLeu	45
QY	241	ACATTGTGCTTGAATGGTCCCTGAAGCGATTCTTTGATCGCTCTCTTCGGAAGATTG	300
DB	46	AlaAlaIrpPheAsnGlyGlyProGlyCysSerSerMetIleGlyLeuPheGlnGluAsn	65
QY	301	GGCCCTTGCAT---GTCAATTCCGACTTTTGATGACATACATCAACCCCTCACTCTCGTGAAC	357
DB	66	GlyProCysHisPheValAsnGlyAspSerThrProSerLeuAsnGluAsnSerIrpAsn	85
QY	358	GAGGTCTCCAATTACTATTCTCTGCCAGCCATGGAGTGGCTTTTCATATAGTCAT	417
DB	86	AsnTyrAlaAsnMetIleTyrIleAspGlnProIleGlyValGlyPheSerTyrGly---	104
QY	418	ACGTTGATGGGTCCATTACCCTGTAACTGGGTGCTCGAAAAATTCAGCTTTCGAGGA	477
DB	104	-----	104
QY	478	GTTTCAGGGCGGTACCCAAACATTGATGCCACTCTGATCGATACACTACCACTTCGCCGA	537
DB	105	-----ThrAspAspValThr-----SerThrValThrAlaAla	115
QY	538	GAGCGCGTGGGAGATCTCTGACAGGATTCCTTACTGGACTACCTAGCTTGGACCTAGG	597
DB	116	ProTyrValIrpAsnLeuLeuGlnAlaPheTyrAlaGlnArgProGluTyrGluSerArg	135
QY	598	GTGCAGTCTAAGGACTTCAGTCTATGGACGAGAGCTATGGAGGCACTATGGTCTGCA	657
DB	136	-----AspPheAlaIlePheIrpGluSerTyrGlyGlyHisTyrGlyProGlu	151
QY	658	TTCTTCAATCATTTTACGACGACAATGAGAGAATGCCACGGTAGTGTAAATGGTGT	717
DB	152	PheAlaSerTyrIleGluGlnAsnAlaAlaIleLysAlaGlySerValThrGlyGln	171
QY	718	CAGCTTAATTCAACTCTCTGGGAATTATTACGGCATCATCGCAGGCGATCCAGGCC	777
DB	172	AsnValAsnIleValAlaLeuGlyValAsnAsnGlyTyrIleAspSerThrIleGlnGlu	191
QY	778	CTTACTACCTGANTTGCCTCTGACAAATACCTACGGTATCAAGGCTGTCAACAGACAC	837
DB	192	LysAlaTyrIleAspPheSerTyrAsnAsnSerTyrGlnGlnIleIleAspSerSerThr	211
QY	838	GTCACAACTACATCAAGTTTGCCACCAAAATGCCAAATGGTGGCCAGGATTTGATTCC	897
DB	212	ArgAspSerLeuLeuAspAlaTyrAsn-----AsnGlnCysLeuProAlaLeuGln	228
QY	898	ACCTGCAACACGACAAACCGCATCGGATTTAGTGTACTACGCCCTCTGCGCCGAGCCACC	957
DB	229	GlnCysSerGlnSerGlySerThr-----SerAspCysThrAsnAlaAsp	243
QY	958	ACATGCTGCAGGACAAATGTTGAGGGCCATACTACGCCCTTTCCTGGTCTGGTGTGTAT	1017
DB	244	SerValCysTyrGlnAsnIleGluGlyProIleSerSerSerGlyAspPheAspValTyr	263

QY	1018	GATATTCGGCATCCATATGATGACCCGACATCCCGCAAGTTATTACACAAATTCCTGGCA	1077
Db	264	AspIleArgGluProSerAsnAspProTyrProPheLeuSerThrTyrLeuSer	283
QY	1078	AAGGACTCTGTCATGAGCGCTATCGGGCTCAACATCACTACACCGAGTCCCATATGAC	1137
Db	284	AspProThrValValLysAlaIleGlyAlaArgThrAsnTyrGlnGluCysProAsnGly	303
QY	1138	GTCCTACTACGGTTTCCAGCAACAGCGGACTTTCTCGGCCCAACTCATCGAAGACCTC	1197
Db	304	ProTyrAsnLysPheAlaSerThrGlyAsnAspPro---ArgSerPheLeuSerThrLeu	322
QY	1198	GAGGAGATCCCTGCTCTCCCGCTCGCGTGTCTCCCTCATCTATGCGACGCGGATACATC	1257
Db	323	SerSerValValGlnSerGlyIleAsnValLeuValTrpAlaGlyAspAlaAspTrpIle	342
QY	1258	TGCAACTGTTGCGCGGTGACGGCGTTTCCTCGCTCGGAACTACTCCCAAGCGGCCAG	1317
Db	343	CysAsnTrpLeuGlyAsnTyrGluValAlaAlaAsnAlaValAspPheProGlyAsnAlaGln	362
QY	1318	TTCGCAAGCGGAGGTACACGCCCTCGAACTCAACGCGTCGAGTATGGGAACTCGC	1377
Db	363	PheSerAlaLeuAspLeuAlaProTyrThrValAsnGlyValGluLysGlyGlnPheLys	382
QY	1378	GAGTATGGTAAATTCCTCTCCTCATCGCTCTATGAGGAGCGCATGAGTCCCACTACTAC	1437
Db	383	ThrValAspAsnPheSerPheLeuLysValTyrGlyAlaGlyHisGluValProTyrTyr	402
QY	1438	CAGCCCATCGCTCCCTGCAATTGTTAAACGGGACTATCTTCGGTGGGATATCGCAGAG	1497
Db	403	GlnProAspThrAlaLeuGlnAlaPheLysGlnIleIle-----	415
QY	1498	GGCCAGAGAAG	1509
Db	416	---GlnLysLys	418

RESULT 2

CBPV_SCHPO STANDARD; PRI: 1002 AA.

ID Q13849; O14366;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Carboxypeptidase Y precursor (EC 3.4.16.5) (CPY).

GN CPY1 OR PCY1 OR SPAC19G12.10C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OX NCBI_TaxID=4896;

[1]

RN SEQUENCE FROM N.A., SUBUNIT, SUBCELLULAR LOCATION, AND MUTAGENESIS OF

RP CYS-627.

RP MEDLINE=97352672; PubMed=9209031;

RA Tabuchi M., Iwihara O., Ohtani Y., Ohuchi N., Sakurai J.-I.,

RA Morita T., Iwihara S., Takegawa K.;

RT "Vacuolar protein sorting in fission yeast: cloning, biosynthesis,

RT transport, and processing of carboxypeptidase Y from

RT Schizosaccharomyces pombe.";

RL J. Bacteriol. 179:4179-4189(1997).

[2]

RN SEQUENCE FROM N.A.

RP STRAIN=972;

RC MEDLINE=21848401; PubMed=11859360;

RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne K., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth N., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mangall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

RESULT 2

CPBPY_SCHPO STANDARD; PRT; 1002 AA.

1D CBPY_SCHPO
AC 013649; 014366;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase Y precursor (EC 3.4.16.5) (CPY).
GN CPY1 OR CPY1 OR SPAC19C12.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RN SEQUENCE FROM N.A., SUBUNIT, SUBCELLULAR LOCATION, AND MUTAGENESIS OF
RP CYS-627.
RX MEDLINE=97352672; PubMed=9200301;
RA Tabuchi M., Iwahara O., Ohtani Y., Ohuchi N., Sakurai J.-I.,
RA Morita T., Iwahara S., Takegawa K.;
RT "vacuolar protein sorting in fission yeast: cloning, biosynthesis,
RT transport, and processing of carboxypeptidase Y from
RT Schizosaccharomyces pombe.";
RL J. Bacteriol. 179:4179-4189(1997).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne K., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

QY 793 TTCGTCGTGGAACAATACCTACGGT-----ATCAAG 822
 Db ||||| :|||
 QY 774 MetAlaCysGluSerProTyrGlyProIleMetSerGlnGluGluCysAspArgIleThr 793
 QY 823 GCTGTCAAGAGACCGCTCTACAACTACATGAAGTTGCGCAACAAATGCAATGGTTGC 882
 Db :|||
 QY 794 GlyAlaTyrAspThr-----Cys 799
 QY 883 CAGGATTTGATTTCCACCTGGAACAGACAAACCCGACCGCATCTAGCTACATCGCCCTC 942
 Db ||||| :|||
 QY 800 AlaLysLeuIleThrGlyCysGlnThrGlyPheThr-----ProVal 814
 QY 943 TCGCGCGAAGCCACCAATCTGTCAGGACAAATGTTGAGGGGCCATACATACGCTTTGCT 1002
 Db ||||| :|||
 QY 815 CysIleGlyAlaSerLeuTyrCysAsnAsnAlaMetIleGlyPro---PheThrLysThr 833
 QY 1003 GPTGCTGGTGTATGATATGTCGATCCATATGATGACCGGACTCGGCCAAGTTAT--- 1059
 Db ||||| :|||
 QY 834 GlyLeuAsnIleTyrAspIleArgGluGluCysArgAspGlnGluHisLeuCysTyrPro 853
 QY 1060 -----TACACAAATTTCTGCGAAGGACTCTGTCATGAGCGTATCGCGCTC 1107
 Db ||||| :|||
 QY 854 GluThrGlyAlaIleGluSerTyrLeuAsnGlnGluPheValGlnGluAlaLeuGlyVal 873
 QY 1108 AACATCAACTACACCCAGTCCAATAATGACTCTACTACGCTTTCCAGCAACAGCGGAC 1167
 Db :||| :|||
 QY 874 GluTyrAspTyrLysGlyCysAsnThrGluValAsnIleGlyPheLeuPheLysGlyAsp 893
 QY 1168 TTGTGTCGCCCACTTCATGCAAGACTCGAGAGATCCTT-----GCTCTCCCGTG 1221
 Db :||| :|||
 QY 894 TrpMetArgLysThrPheArgAspValThrAlaIleLeuGluAlaGlyLeuProVal 913
 QY 1222 CGTGTCTCCCTCATCTAT---GGCGAGCGGATTCATCTGCAACTGCTCGGGGGTCCAG 1278
 Db ||||| :|||
 QY 914 -----LeuIleTyrAlaGlyAspAlaAspTyrIleCysAsnTyrMetGlyAsnGlu 930
 QY 1279 GCGGTTTCCCTGCTGCGAACTACTCCAGCCCGCAGTTCGGAAGCGCAGGATACAG 1338
 Db ||||| :|||
 QY 931 AlaTrpThrAspAlaLeuGluTyrPalaGlyGlnArgGluPheTyrGluAlaGluLeuLys 950
 QY 1339 CCCCTGAAGTCAACGCGCTCGATGATGGGAAACTCCGAGATGTTGTAATTTCTCCTTC 1398
 Db ||||| :|||
 QY 951 ProTrpSerProAsnGlyLysGluAlaGlyArgGlyLysSerPheLysAsnPheGlyTyr 970
 QY 1399 ACTCGGCTCATGAGGAGGATCACTCCATCCATACAGCCATCGCTCCCTCGAA 1458
 Db ||||| :|||
 QY 971 LeuArgLeuTyrGluAlaGlyHisMetValProPheAsnGlnProGluAlaSerLeuGlu 990
 QY 1459 TTGTTTAAACGCGACTATCTCGGT 1482
 Db :||| :|||
 QY 991 MetLeuAsnSerTrpIleAspGly 998
 RESULT 3
 ID YBY9_YEAST STANDARD; PRT; 508 AA.
 AC P38109.
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative serine carboxypeptidase in BSRI-IRAI intergenic region
 DE (EC 3.4.16.-)
 GN YBRI39W OR YBRI1015.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=288c;
 RX MEDLINE=9437817; PubMed=8091856;
 RA Becam A.-M., Cullin C., Grzybowska E., Lacroite F., Nasr F.,
 RA Ozier-Kalogeropoulos O., Palucha A., Slonimski P.P., Zagulski M.,
 RA Herbert C.J.;

"The sequence of 29.7 kb from the right arm of chromosome II reveals
 13 complete open reading frames, of which ten correspond to new
 genes.";
 Yeast 10:S1-S11(1994).
 [2]
 DISCUSSION OF SEQUENCE.
 MEDLINE=95042830; PubMed=7954890;
 RA Nasr F., Becam A.-M., Grzybowska E., Zagulski M., Slonimski P.P.,
 RA Herbert C.J.;
 RA "An analysis of the sequence of part of the right arm of chromosome
 II of *S. cerevisiae* reveals new genes encoding an amino-acid permease
 and a carboxypeptidase.";
 Cur. Genet. 26:1-7(1994).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL; X75891; CAA53497.1; -;
 EMBL; Z36008; CAA85097.1; -;
 DR PIR; S46008; S46008.
 DR HSSP; P00729; 1CPI.
 DR MEROPS; S10.0PW; -;
 DR SGD; S0000343; YBRI39W.
 DR InterPro; IPR003179; Ser_estrs_site.
 DR InterPro; IPR001563; Serine_carbpept.
 DR Pfam; PF00450; serine_carbpept; 1.
 DR PRINTS; PR00724; CRBOXYPTASEC.
 DR ProDom; PD001189; Serine_carbpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_HIS; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hypothetical protein; Hydrolase; Carboxypeptidase.
 FT ACT_SITE 219 219 BY SIMILARITY.
 FT ACT_SITE 415 415 BY SIMILARITY.
 FT ACT_SITE 474 474 BY SIMILARITY.
 SQ SEQUENCE 508 AA: 57639 MW: AAB2806C8EE2EDE1 CRC64;
 Alignment Scores:
 Pred. No.: 1.03e-36 Length: 508
 Score: 636.00 Matches: 167
 Percent Similarity: 46.23% Conservative: 70
 Best Local Similarity: 32.62% Mismatches: 197
 Query Match: 21.90% Indels: 78
 DB: 1 Gaps: 14
 US-09-712-338-1-copy_55_1662 (1-1608) x YBY9_YEAST (1-508)
 QY 1 CTTCAGGAAGTACACCGCGCTCGTGGTAGACACAGCTA-----CCCAAG 48
 Db ||||| :|||
 QY 43 LeuProGlnAsnThrGlnGlnThrLeuLysLeuAspArgLeuAsnHisAspProLeu 62
 QY 49 AACCCACCGGGGTCAAGACTCTTACACCGCAACAATGTCACCTCGGTACAGGAA 108
 Db ||| :|||
 QY 63 PheThrThrPheIleSerSerValAspThrAspTyrSerLeuArgLeuArgThrValAsp 82
 QY 109 CCCGGGCGAGAGGGGCTGCGAGACTACCCCGGGTGTCAAACTCTTCTGATATGTC 168
 Db ||||| :|||
 QY 83 ProSerLysLeuGlyIle-----AspThrValLysGlnTrpSerGlyTyrMet 98
 QY 169 GACACCTCTCCGAGTCCCATACCTCTCTGTTGGTTCTTCGAGCCAGACATACCCAGAA 228
 Db ||||| :|||
 QY 99 AspTyrLysAspSerLysHisPhePheTyrTrpPhePheGluSerArgAsnAspProAla 118
 QY 229 ACTGCACCTATCATATGTTGTTGAATGTTGGTGGCCCTGGAGCGGATTTCTTGTGCTC 288
 Db ||||| :|||
 QY 119 AsnAspProIleLeuLeuLysLeuAsnGlyGlyProGlyCysSerPheThrGlyLeu 138
 QY 289 TTCGAGAGTGTGGCGCTCCCTGCCATGTCAATTCGACTTTTGTGATGACTACATCAACCTCAC 348

Db 139 LeuPheGluLeuGlyProSerSerIleGlyAlaAspMetLysProIleHisAsnProTyr 158
 QY 349 TCGTGGACAGGAGTCCCAATTTACTATTCTCTCCAGCCATCGGAGTGGCTTTTCA 408
 Db 159 SerTrpAsnAsnAlaSerMetIlePheLeuGluGlnProLeuGlyValGlyPheSer 178
 QY 409 TATAGTATAGCGTGTGATGGTCCATTAAACCTGTAACTGGGGTCGTCGAAATTCGAGC 468
 Db 179 TyrGlyAsp 181
 QY 469 TTTCAGAGTTCAGGCGCGGTACCCCAACCATTCGATGCCACTCGATCGATACCAAT 528
 Db 182 188
 QY 529 CTTGCCGAGAGCGCGTTCGGAGATCTGCAAGGATTCCTTAGTGCATACCTAGCTG 588
 Db 189 LeuAlaGlyLysAspAlaTyrIlePheLeuGluLeuPheGluAlaPheProHisLeu 208
 QY 589 GACTCTAGGTGAGTCTAAGGACTTCAAGTCTATGAGGAGAGCTATGGAGGCACTAT 648
 Db 209 224
 QY 649 GGTCTGCACTTCCTCAATCATTTTACGACCAAT---GAGAGAAATGCCAGGTAGT 705
 Db 225 IleProGlnIleAlaHisGluIleValValLysAsnProGluArg 239
 QY 706 GTTAATGGTTCAGGCTTAATTCCTCACTCTCTGGGAATTAACGSCATCATCGAGC 765
 Db 240 255
 QY 766 GCGATCCAGCGCCCTTACTACCTCGAATTCGCTGTG---AACAAATACCTACGATCAAG 822
 Db 256 LeuIleGlnAlaAspTyrTrpGluProMetAlaCysGlyLysGlyTyrHisProVal 275
 QY 823 CGTGTCAAGGAGACCGCTACACTACATCAAGTTTGGCAACCAATGCCAAATGGTTC 882
 Db 276 LeuSerSerGluGluCysGluLysMetSerLysAlaAlaGlyArg 290
 QY 883 CAGGATTTGATTCCACCTGCAACACAGAACCGCGCATTCAGTACGACGCTC 942
 Db 291 304
 QY 943 ---TGCGCGGACCGACCAACAGTGCAGGACAATGTTGAGGGCGCATCTACGCG 996
 Db 305 LeuProCysIleValAlaThrAlaTyrCysAspSerAlaLeuLeuGluProTyrIleAsn 324
 QY 997 TTTCGTGGTGGTGTGTATCATATTCGGCATCCATATGATGACCGC 1044
 Db 325 ---ThrGlyLeuAsnValTyrAspIleArgGlyProCysGluAspAsnSerThrAspGly 343
 QY 1045 ---ACTCCGCCAAGTTATTACACAAATTTCTGGCAAGGACTCTCTCATGGAC 1095
 Db 344 MetCysTyrThrGlyLeuArgTyrValAspGlnTyrMetAsnPheProGluValGlnGlu 363
 QY 1096 GCTATCGCGCTCAACATC---AATACACCCAGTCCCAATATGAGTCTACTACGCTTC 1152
 Db 364 ThrLeuGlySerAspValHisAsnTyrSerGlyCysAspAsnAspValPheThrGlyPhe 383
 QY 1153 CAGCAACAGCGGACTTTGTCTGGCCCACTTCATCGAAGACCTCGAGAGATCCTTGT 1212
 Db 384 LeuPheThrGlyAspGlySerLysPro---PheGlnGlnTyrIleAlaGluLeuLeuAsn 402
 QY 1213 CTCCCGTGGCTGCTCCCTCATCTATGCGGACCGGATTCATCTGCAACTGGTTCGGC 1272
 Db 403 HisAsnIleProValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGly 422
 QY 1273 GGTACGCGCTTCCCTCGCTCGGCACTACCCCAAGCGCGGCTCCGAGCGCACGG 1332
 Db 423 AsnHisAlaTrpSerAsnGluLeuGluTrpIleAsnLysArgArgTyrGlnArgArgMet 442
 QY 1333 TACAGCGCC-----CTGAAAGTCAACGGCTCGAGTATGGGAACCTCGCGAGTATGGT 1386

Db 443 LeuArgProTrpValSerLysGluThrGlyGluGluLeuGlyValLysAsnTyrGly 462
 QY 1387 AATTTCTCTTCACTCGCTTATGAGCGCCATGAAGTCCCTACTACCAAGCCCATC 1446
 Db 463 ProPheThrPheLeuArgIleTyrAspAlaGlyHisMetValProTyrAspGlnProGlu 482
 QY 1447 GCCTCCCTGCAATGTGTTAACCGGACTATCTCGGT 1482
 Db 483 AlaSerLeuGluMetValAsnSerTrpIleSerGly 494
 RESULT 4
 ID CBPY_CANAL STANDARD; PRT; 542 AA.
 AC P30574;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).
 GN CPV1
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93051356; PubMed=1427093;
 RA Mukhtar M., Logan D.A., Kauffer N.F.;
 RT "The carboxypeptidase Y-encoding gene from Candida albicans and its
 RL transcription during yeast-to-hyphae conversion.";
 RL Gene 121:173-177(1992).
 CC -1- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES.
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity.
 CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
 CC -1- INDUCTION: TRANSIENTLY DOWN-REGULATED DURING THE EARLY EVENTS OF
 CC YEAST TO HYPAE CONVERSION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M95182; AAA34326.2; -
 CC HSSP; P00729; 1CPY. -
 CC MEROPS; S10.001; -
 CC InterPro; IPR000379; Ser_estrs_site.
 CC InterPro; IPR001563; Serine_carbpept.
 CC Pfam; PF00450; serine_carbpept; 1.
 CC PRINTS; PR00724; CRBOXYPTASEC.
 CC ProDom; PD001189; Serine_carbpept; 1.
 CC PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 CC PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 CC Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
 KW SIGNAL
 FT PROPEP 1 21
 FT CHAIN 127
 FT ACT_SITE 128 542
 FT ACT_SITE 269 269
 FT ACT_SITE 461 461
 FT ACT_SITE 518 518
 FT BINDING 519 519
 FT BINDING 464 464
 FT DISULFID 182 421
 FT DISULFID 316 330
 FT DISULFID 340 363
 FT DISULFID 347 356
 FT DISULFID 385 391
 FT CARBOHYD 213 213
 FT CARBOHYD 291 291
 FT SIGNAL 1 21
 FT PROPEP 22 127
 FT CHAIN 128 542
 FT ACT_SITE 269 269
 FT ACT_SITE 461 461
 FT ACT_SITE 518 518
 FT BINDING 519 519
 FT BINDING 464 464
 FT DISULFID 182 421
 FT DISULFID 316 330
 FT DISULFID 340 363
 FT DISULFID 347 356
 FT DISULFID 385 391
 FT CARBOHYD 213 213
 FT CARBOHYD 291 291
 SEQUENCE 542 AA; 61044 MW; 7FA6B9F82F9D44F CRC64;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DB EMBL; X87987; CAA61240.1; -
 DR PIR; S61713; S61713.
 DR HSSP; P00729; IYSC.
 DR MEROPS; S10.001; -
 DR InterPro; IPR000379; Ser_estrs_site.
 DR InterPro; IPR001563; Serine_carbpept.
 DR Pfam; PF00450; serine_carbpept; 1.
 DR PRINTS; PR00724; CRBOXPTASEC.
 DR ProDom; PD001189; Serine_carbpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 20
 FT PROPEP 21 107
 FT CHAIN 108 523
 FT ACT_SITE 249 249
 FT ACT_SITE 441 441
 FT ACT_SITE 498 498
 FT DISULFID 162 401
 FT DISULFID 296 310
 FT DISULFID 320 343
 FT DISULFID 327 336
 FT DISULFID 365 371
 FT CARBOHYD 193 193
 FT CARBOHYD 271 271
 FT CARBOHYD 484 484
 FT CARBOHYD 487 487
 SQ SEQUENCE 523 AA; 59447 MW; CA6093BDE2E9D178 CRC64;
 Alignment Scores:
 Pred. No.: 8.08e-34
 Score: 594.50
 Percent Similarity: 45.89%
 Best Local Similarity: 33.12%
 Query Match: 20.47%
 DB: 1
 Gaps: 16

US-09-712-338-1_COPY_55_1662 (1-1608) x CBPY_PICPA (1-523)

QY	148	AAATCTACTCTGGATATGTCGACAC	---	TTCCCGAGTCCCATACCTTCTTCIGTTC	204
Db	121	LysGlnTyrSerGlyTyrLeuAspValGluAlaAspAspLysHisPheTyrTrpPhe	140		
QY	205	TTGGAAGCCAGACATAACCCAGAAACTGCACCTATCACATTGGTGGTGAATGGTGGCCCT	264		
Db	141	PheGluSerArgAsnAspProGlnAsnAspProIleLeuTrpLeuAsnGlyGlyPro	160		
QY	265	GGACCGCATCTTTGATCGGTCTCTTCGAGAGTTCGGGCCCTGCCATGTCATTCGACT	324		
Db	161	GlyCysSerSerLeuThrGlyLeuPhePheGluLeuGlySerSerArgIleAsnGluAsn	180		
QY	325	TTTGATGACTACATCAACCTCACCTGCGAGAGGTCTCCAAITTAATTCCTGTC	384		
Db	181	LeuLysProIlePheAsnProTyrSerTrpAsnGlyAsnAlaSerIleIleTyrLeuAsp	200		
QY	385	CAGCCATTCGGGTGGCTTTTCATATAGTATGATACGGTTCATGGGTCATTAACCTGTGA	444		
Db	201	GlnProValAsnValGlyPheSerTyrSer	210		
QY	445	ACIGGGGTGCGAAATTCGAGCTTTGACGAGGTTCAGGGCCGGTACCCACCATIGAT	504		
Db	210				
QY	505	GCCACTCTCATGACTACCAATCTGCCCGCAGAGGGCGGTGGGAGATCCGCAAGCA	564		
				

Db	211	SerSerSerValSerAsnThrValValAlaGlyGluAspValTyrAlaPheLeuGlnLeu	230
QY	565	TTTCCTTAGTGGACTAAGTACTAGTGGACTCTAGGCTCAGTCTAAGGACTTCAGTCTATGG	624
Db	231	PhePheGlnHisPhePro	246
QY	625	ACGGAGAGCATGGAGGGCACTATGGCTCTGATCTTCAATCATTTTACGAGCAGAAAT	684
Db	247	GlyGluSerTyrAlaGlyHisTyrIleProValPheAlaAspGluLeuLeuSerGlnLys	266
QY	685	GAGAGAAATGGCCAAAGGTAGTGTAAATGGTGTTCAGCTTAATTTCACTCTCTGGAAT	744
Db	267	AsnArg	275
QY	745	ATT-----AACGGCATCATCAGCAGGCGATCCAGGCCCTTACTTACCTGAATTCGT	798
Db	276	LeuIleGlyAsnGlyLeuThrAspProLeuThrGlnTyrArgTyrIleGluProMetAla	295
QY	799	GTGAACAATACCTACGGTATCAAGGCTGTCACGAGACCGCTCTACAACTACATGAAT	858
Db	296	CysGlyGluGly--GlyAlaProSerValLeuPro	306
QY	859	CCCAACCAATGCCAAAT-----GGTGGCAGATTTGATTTCACCC	900
Db	307	AlaAspGluCysGluAsnMetLeuValThrGlnAspLysCysLeuSerLeuIleGlnAla	326
QY	901	TGCAAAACAGACAAACCGCACCGCATAGCTACTACGCCCTCTCGCGGAGGACCAAC	960
Db	327	CysTyrAspSerGlnSerAla-----PheThrCysAlaProAlaAlaIle	341
QY	961	ATGTGACGGGACATGTTGAGGGGCGCATACACGCCITTTGCTGTCGTGTGTATGAT	1020
Db	342	TyrCysAsnAsnAlaGlnMetGlyProTyrGlnArg---ThrGlyLysAsnValTyrAsp	360
QY	1021	ATTCGGCATCCATATGATGACCCGACT-----CGCCCAAGTATTACAAACA	1068
Db	361	IleArgLysGluCysAspGlyGlySerLeuCysTyrLysAspLeuGluPheIleAspThr	380
QY	1069	TTTCGTGCAAAAGGACTCTGTCATGGAGCTATCGCGGTCAACATCAAC---TACACCCAG	1125
Db	381	TyrLeuAsnGlnLysPheValGlnAspAlaLeuGlyAlaGluValAspThrTyrGluSer	400
QY	1126	TCCAATAATGACGTCTACTAGGCTTCCAGCAACAGCGGACITTTGCTGCGCCCACTTC	1185
Db	401	CysAsnPheGluIleAsnArgAsnPheLeuPheAlaGlyAspTrpMetLysPro---Tyr	419
QY	1186	ATCGAAGCACTCGAGGAGATCCCTT-----GCTCTCCCGTGGCTGTCTCCCTCATCTAT	1239
Db	420	HisGluHisValSerSerLeuLeuAsnLysGlyLeuProVal-----LeuIleTyr	436
QY	1240	---GGCGACCCGATTACATCTCACTGCTCGCGGTCCGGCTCAGGCGCTTCCCTCGCTGG	1296
Db	437	AlaGlyAspLysAspPheIleCysAsnTrpLeuGlyAsnArgAlaTrpThrAspValLeu	456
QY	1297	AACTACTCCCAACCGGCCCTTCGAGGAGGTTCACGCGCCCTGGAAGTCAACGGC	1356
Db	457	ProTrpValAspAlaAspGlyPheGluLysAlaGluValGlnAspTrpLeuValAsnGly	476
QY	1357	GTGAGTATGGGAAACTCGGAGTATGTAATTTCTCTTCTACTCGCGTCTATGAGGCA	1416
Db	477	ArgLysAlaGlyGlyPheLysAsnTyrSerAsnPheThrTyrLeuArgValTyrAspAla	496
QY	1417	GGCCATGAAGTCCCATACACAGCCCATCGCTCCCTGCAATTTGTTACCGGACATATC	1476
Db	497	GlyHisMetAlaProTyrAspGlnProGluAsnSerHisGluMetValAsnArgTrpIle	516
QY	1477	TTCCGGT 1482	
Db	517	SerGly 518	

RESULT 6

CBPY_YEAST

ID CBPY_YEAST

STANDARD;

FRT;

532 AA.

AC	P00729;	CC	GOLGI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM
DT	21-JUL-1986 (Rel. 01, Created)	CC	SECRETORY PROTEINS AT OR BEYOND A LAIE GOLGI COMPARTMENT AND IS
DT	01-NOV-1988 (Rel. 09, Last sequence update)	CC	SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR
DT	15-SEP-2003 (Rel. 42, Last annotation update)	CC	ENDOSOME-LIKE COMPARTMENT. UPON ARRIVAL IN THE VACUOLE, THE
DE	Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).	CC	N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMATICAL
GN	PRC1 OR YMR297W.	CC	ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).
OS	Saccharomyces cerevisiae (baker's yeast).	CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	CC	-!- DATABASE: NAME-Worthington enzyme manual;
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.	CC	WWW="http://www.worthington-biochem.com/manual/C/COY.html".
OX	NCBI_TaxID=4932;	CC	-----
RN	[1]	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
RP	SEQUENCE FROM N.A.	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
RX	MEDLINE=87131100; PubMed=3028649;	CC	the European Bioinformatics Institute. There are no restrictions on its
RA	Valls L.A., Hunter C.P., Rothman J.H., Stevens T.H.;	CC	use by non-profit institutions as long as its content is in no way
RA	"Protein sorting in yeast: the localization determinant of yeast	CC	modified and this statement is not removed. Usage by and for commercial
RT	vacuolar carboxypeptidase Y resides in the propeptide."	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
RL	Cell 48:887-897(1987).	CC	or send an email to license@isb-sib.ch).
RL	[2]	CC	-----
RP	Nature 387:90-93(1997).	DR	EMBL; M15482; AAA34902.1; -.
RP	[3]	DR	EMBL; X80836; CAA56806.1; -.
RC	SEQUENCE FROM N.A.	DR	PIR; A26597; CPBY.
RX	STRAN=5288c / AB972;	DR	PDB; 1YSC; 22-JUN-94.
RA	PubMed=9169872;	DR	PDB; 1CPY; 15-SEP-95.
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,	DR	MROPS; S10.001; -.
RA	Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,	DR	SGD; S0004912; PRC1.
RA	Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,	DR	GO; GO:0005783; C:endoplasmic reticulum; IDA.
RA	Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;	DR	InterPro; IPR000379; Ser_estrs_site.
RT	"the nucleotide sequence of Saccharomyces cerevisiae chromosome	DR	Pfam; PF00450; serine_carbpept; 1.
RT	XIII."	DR	PRINTS; PR00724; CRBOXYPTASEC.
RL	[4]	DR	Prodom; PD001189; Serine_carbpept; 1.
RP	Nature 387:90-93(1997).	DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
RP	[5]	DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
RA	Sevendsen I., Martin B.M., Viswanatha T., Johansen J.T.;	DR	Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal;
RA	"Amino acid sequence of carboxypeptidase Y. II. Peptides from	KW	3D-structure.
RT	enzymatic cleavages."	KW	POTENTIAL.
RL	Carlsberg Res. Commun. 47:15-27(1982).	FT	SIGNAL
RL	[6]	FT	PROPEP
RP	REVISIONS, AND ACTIVE SITE SER-257.	FT	CHAIN
RA	Braddam K., Svendsen I.;	FT	ACT_SITE
RT	"Identification of methionyl and cysteinyl residues in the substrate	FT	ACT_SITE
RT	binding site of carboxypeptidase Y."	FT	ACT_SITE
RL	Carlsberg Res. Commun. 49:639-645(1984).	FT	BINDING
RP	[7]	FT	BINDING
RP	ACTIVE SITE HIS-508.	FT	DISULFID
RA	Bech L.M., Braddam K.;	FT	DISULFID
RT	"Inactivation of carboxypeptidase Y by mutational removal of the	FT	DISULFID
RT	putative essential histidyl residue"	FT	DISULFID
RL	Carlsberg Res. Commun. 54:165-171(1989).	FT	CARBOHYD
RN	[8]	FT	CARBOHYD
RP	MUTAGENESIS.	FT	CARBOHYD
RP	MEDLINE=94114535; PubMed=7904479;	FT	CARBOHYD
RA	Mortensen U.H., Remington S.J., Braddam K.;	FT	CARBOHYD
RT	"Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen	FT	CARBOHYD
RT	bond network stabilizes the transition state by interaction with the	FT	CARBOHYD
RT	C-terminal carboxylate group of the substrate."	FT	CARBOHYD
RL	Biochemistry 33:508-517(1994).	FT	SITE
RN	[9]	FT	MUTAGEN
RP	X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).	FT	CONFLICT
RX	MEDLINE=95244421; PubMed=7727362;	FT	CONFLICT
RA	Endrizzi J.A., Braddam K., Remington S.J.;	FT	CONFLICT
RT	"2.8-A structure of yeast serine carboxypeptidase."	FT	STRAND
RL	Biochemistry 33:11106-11120(1994).	FT	STRAND
CC	-!- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS	FT	TURN
CC	PREFERENTIALLY PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC	FT	TURN
CC	RESIDUE IN P1' POSITION, AS WELL AS METHIONINE, LEUCINE OR	FT	TURN
CC	PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.	FT	TURN
CC	-!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a	FT	TURN
CC	broad specificity.	FT	HELEX
CC	-!- ENZYME REGULATION: INHIBITED BY ZPECK.	FT	TURN
CC	-!- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.	FT	TURN
CC	-!- PTM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZYMOGEN AND	FT	TURN
CC	IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO	FT	TURN
CC	A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE	FT	TURN
CC	GOLGI, EXTENSION OF ITS CORE OLIGOSACCHARIDES LEADS TO THE	FT	TURN

QY 1465 AACCGG 1470
 Db 489 LysArg 490
 RESULT 8
 CBPX_ORYSA
 ID CBPX_ORYSA STANDARD; PRT; 429 AA.
 AC P52712;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Serine carboxypeptidase-like precursor (EC 3.4.16.-).
 GN CBP31.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Pharbitidae; Oryzoideae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RF SEQUENCE FROM N.A.
 RC STRAIN=cv. Yukihikari;
 RA Washio K., Ishikawa K.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
 CC !- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity.
 CC !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D17587; BA045111.1; -
 DR PIR: T03607; T03607.
 DR HSP: P00729; LYSC.
 DR MEROPS: S10.009; -
 DR Gramene; P52712; -
 DR InterPro: IPR000379; Ser_estrs_site.
 DR InterPro: IPR001563; Serine_carbpept.
 DR Pfam: PF00450; serine_carbpept; 1.
 DR PRINTS: PR00724; CRBOXPTASEC.
 DR ProDom; P0001189; Serine_carbpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Signal.
 FT SIGNAL 1 ?
 FT CHAIN ? 429
 FT ACT_SITE 148 148
 FT ACT_SITE 336 336
 FT ACT_SITE 393 393
 FT BINDING 339 339
 FT CARBOHYD 76 76
 FT CARBOHYD 414 414
 FT CARBOHYD 417 417
 SQ SEQUENCE 429 AA; 47746 MW; 1D5A668544325BB1 CRC64;
 Alignment Scores:
 Pred. No.: 2,23e-25 Length: 429
 Score: 473.00 Matches: 130
 Percent Similarity: 45.49% Conservative: 77
 Best Local Similarity: 28.57% Mismatches: 178
 Query Match: 16.29% Indels: 70
 DB: 1 Gaps: 16
 US-09-712-338-1_COPY_55_1662 (1-1608) x CBPX_ORYSA (1-429)
 QY 154 TACTICNGGATAT-----GTCGACACCTCCCGAGTCCCATACCTCTTCGTTCTTC 207
 Db 17 TACTICNGGATAT-----GTCGACACCTCCCGAGTCCCATACCTCTTCGTTCTTC 207

FT BINDING 407 SUBSTRATE (BY SIMILARITY).
 FT CARBOHYD 144 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 500 AA; 55446 MW; AE455E2780147DB8 CRC64;

Alignment Scores:
 Pred. No.: 6,79e-24 Length: 500
 Score: 452.00 Matches: 133
 Percent Similarity: 43.56% Conservative: 80
 Best local Similarity: 27.20% Mismatches: 206
 Query Match: 15.56% Indels: 70
 DB: 1 Gaps: 17

US-09-712-338-1_COPY_55_1662 (1-1608) x CBP3_ORYSA (1-500)

QY	40	CTACCCAAAG-----AACCCACACGGG-----GTCACAGCTCTTACACCCCAAC	84
Db	47	LeuProLysGluAlaGlyProThrGlyAlaGlyAspValProSerValAlaProGlyGlu	66
QY	85	AATGTCACCAACCGGTACAGGAACCCGGG---GCAGAGGGCGTTCGGAGACTACCCCG	141
Db	67	LeuLeuGluAArgValThrLeuProGlyLeuProGlyValGlyAsp-----	83
QY	142	GGTGTCAAATCCTACTCGGATAT-----GTGACACCTCTCCGAGTCCCATACCTTC	195
Db	84	---LeuGlyHisHisAlaGlyTyrArgLeuProAsnThrHisAspAlaArgMetPhe	102
QY	196	TTCGTGTTCTTCGAAGCAGACATAACCCAGAACTGCACCTATACATTCGTTGTAAT	255
Db	103	TyrPheLeuPheGluSerArgGlyLysGluAsp---ProValValIleThrPheThr	121
QY	256	GGTGGCCCTGGAACCGATCTTTCGTCGTCCTTCGAAGAGTTGGGCCCTTCGATGTC	315
Db	122	GlyGlyProGlyLysSerSerGluLeuAlaValPheTyrGluAsnGlyPropheThrIle	141
QY	316	AATTCGACTTTTGATGACTACATCAACCCCTACTCGTGGACAGAGGTCICCAATTTACTA	375
Db	142	SerAsnAsnMetSerLeuAlaIleAsnLysPheGlyTyrAspThrIleSerAsnIleIle	161
QY	376	TTCGTGTCGCCGACCATGGAGTCGGCTTTTCATATAGTATGATACGGTTGATGGTCCATT	435
Db	162	PheValAspGlnProThrGlyThrGlyPheSerTyrSerAspAspThrArg	181
QY	436	AACCTGTACTGGGTCGTCGAAATTCAGGCTTTCAGGAGTTCAGGCGCGGTACCCA	495
Db	182	HisAspGluThrGlyVal-----	187
QY	496	ACCATTCATCCACTCTGATCGATACCTACCAATCTTCCCGACAGGCGGCTTGGAGATC	555
Db	188	-----SerAsnAspLeuTyrSerPhe	194
QY	556	CTCAGAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGTCGAGTCTAAGGACTTC	615
Db	195	LeuGluValPhePheLysLysHisProGluPhe-----AlaLysAsnAspPhe	210
QY	616	AGTCTATGACGAGACGATNAGGAGGCACACTATGGTCCTCAITCTTCAATCATTTTAC	675
Db	211	PheIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaSerArgValHis	230
QY	676	GAGCACAATCAGAGAATGCCACGCTAGTCTTAATGGTTTCAGCTTAATTTCACTCT	735
Db	231	GlnGlyAsnLys-----AlaAsn-----GluGlyIleHisIleAsnLeuLysGly	245
QY	736	CTGGGAATTAATAACGGCATCATCGACGAGGCGATCCAGCCCTTACTACCTGATTC	795
Db	246	PheAlaIleGlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaIleThrAspTyr	265
QY	796	CGTGTGAACATACCTACGGTATCAAGGCTGTCACAGCAGACCGCTTACAACTACATGAAG	855
Db	266	AlaLeuAsp-----MetAsnLeuIleLysSerAspTyrAspArgIle---	280
QY	856	TTTGCAACCAATGCCAATGTTGCCAGGATTTGATTTCCACCTGCAACAGACAAAC	915
Db	281	-----AsnLysPheIleProCysGluPheAlaIleLysLeuLysGlyThrAsnGly	298

QY	916	CCACGCCGATTAGCTGACTACGCCCTCTCGCGGAAGCCACCAACATGTGCGAGGACAAT	975
Db	299	Lys-----AlaSerCysMetAlaAlaTyrMetValCysAsnSerIle	312
QY	976	GTITAGGGGCCATACCTACCCCTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1035
Db	313	PheSerSerIleMetLysLeuValGlyIleThrLysAsnTyrTyrAspValArgLysGluCys	332
QY	1036	GATGACCCGACTCCGCCAAGTATTACAAC-----AAATTTCTGCCAAGACTCTGTC	1089
Db	333	GluGlyLysLeuCysTyrAspPheSerAsnLeuGluLysPhePheLysAspLysAlaVal	352
QY	1090	ATFGACGCTAICGGGCTC---AACATCAACTACACCCACTCCCAATAATCACTCTACTAC	1146
Db	353	LysGluAlaIleGlyValGlyAspLeuGluPheValSerCysSerThrThrValTyrGln	372
QY	1147	GCTTTCCAGCAACAGGCGACTTGTCTGGCCCACTTCATCGAA---GACCTCGAGGAG	1203
Db	373	AlaMetLeuThrAsp-----TrpMetArgAsnLeuGluValGlyIleProAla	388
QY	1204	ATCCCTGTCTCTCCCGTGGTGTCTCCCTCACTATGTCGCGGCGCGGATACATCTGCAAC	1263
Db	389	LeuLeuGluAspGlyIleAsnValLeuIleTyrAlaGlyGluTyrAspLeuIleCysAsn	408
QY	1264	TGGTTCGCGGTCAGGCGCTTTCCTCGCTGCGAAGTCTCCCAAGCCGCCAGTCCGA	1323
Db	409	TrpLeuGlyAsnSerArgTyrValHisSerMetGluTyrSerGlyGlnLysAspPheVal	428
QY	1324	AGCGCAGGTACACGCCCTGAACTCAAGCGCTGCGAGTATGCGGGAACCTCGCGAGTAT	1383
Db	429	SerSerHisGluSerProPheValValAspGlyAlaGluAlaGlyValLeuLysSerHis	448
QY	1384	GGAATTTCTCTCTCCTCCTCTATGAGCAGCGCCATGAATCCCATCTACCTACACGCC	1443
Db	449	GlyProLeuSerPheLeuLysValHisAsnAlaGlyHisMetValProMetAspGlnPro	468
QY	1444	ATCGCTCCCTCGCAATGTTTAAACGGG	1470
Db	469	LysAlaSerLeuGluMetLeuArgArg	477

RESULT 10
 CBP3_HORVU STANDARD; PRT; 508 AA.
 ID CBP3_HORVU AC P21529;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Serine carboxypeptidase III precursor (BC 3.4.16.5) (CP-MIII).
 GN CBP3 OR EXP;3.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Himalaya; TISSUE=Aleucone;
 RA Rocher A., Lok F., Cameron-Mills V., von Wettstein D.:
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 81-491.
 RC STRAIN=cv. Gula;
 RX MEDLINE=90315015; PubMed=2639682;
 RA Soerensen S.B., Svendsen I., Breddam K.:
 RI "Primary structure of carboxypeptidase III from malted barley.";
 RL Carlsberg Res. Commun. 54:193-202(1989).
 CC -I- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity.
 CC -I- ENZYME REGULATION: INHIBITED BY MERCURIC IONS.
 CC -I- SUBUNIT: Monomer.
 CC -I- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOSPERM.
 CC -I- DEVELOPMENTAL STAGE: EXPRESSED MAINLY IN THE ALBURONE AND, TO A

CC LESSER EXTENT IN THE EMBRYO, THROUGHOUT THE 5-DAYS GERMINATION
 CC PERIOD EXCLUSIVELY, WITH A MAXIMAL LEVEL AT 3 DAYS. ALSO FOUND IN
 CC THE ROOTS AND SHOOTS OF THE GROWING SEEDLING.
 CC -I- SIMILARITY: BELONGS TO PETIDASE FAMILY S10.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Y09604; CAA70817.1; ..
 DR HSP; P00729; ICPY; ..
 DR MEROPS; S10.009; ..
 DR InterPro; IPR000379; Ser_estrs_site.
 DR InterPro; IPR001563; Serine_carbpept.
 DR Pfam; PF00450; serine_carbpept; 1.
 DR PRINTS; PR00724; CRBOXYPASEC.
 DR ProDom; PD001189; Serine_carbpept; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER; 1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 DR Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
 KW SIGNAL 1 19
 FT PROPEP 20 80
 FT CHAIN 81 491
 FT PROPEP 492 508
 FT MOD_RES 81 81
 FT ACT_SITE 223 223
 FT ACT_SITE 411 411
 FT ACT_SITE 468 468
 FT BINDING 414 414
 FT CARBOHYD 151 151
 FT VARIANT 265 265
 SQ SEQUENCE 508 AA; 56362 MW; 70C6751D78D40AB6 CRC64;

Alignment Scores:

Pred. No.: 1,23e-22 Length: 508
 Score: 434.00 Matches: 132
 Percent Similarity: 41.97% Conservatives: 77
 Best local Similarity: 26.51% Mismatches: 221
 Query Match: 14.94% Indels: 68
 Db: 1 Gaps: 15

US-09-712-338-1_copy_55_1662 (1-1608) x CBP3_HORVU (1-508)

QY 1 CTTCCAGGAGTACACCGGCTCGTCCGTACAGACAGCTACCCAGAACCCACCGGG 60
 Db 47 LeuProLysAspSerSerSerSerGlyArgHisGly---AlaArgValGlyGluGly 65
 QY 61 GTCAAGACTCTTACACCGCAACCAATCTCACCTCCGTACAGGAACCGGG---GCA 117
 Db 56 AsnGluAspValAlaProGlyGlnLeuLeuGluArgValThrLeuProGlyLeuPro 85
 QY 118 GAGGCGCTCGCAGACTACCCGGGTGTCAAATCCTACTCTGGATAT-----GTGAC 171
 Db 86 GluGlyValAlaAsp-----LeuGlyHisHisAlaGlyTyrTyrArgLeuPro 101
 QY 172 ACCTCTCCGAGTCCCATACCTCTCTCTGCTTCCTCGAGCCAGCATACCCAGAACT 231
 Db 102 AsnThrHisAspAlaArgMetPheTyrPhePheGluSerArgGlyLysGluAsp 121
 QY 232 GCACATATCATATTTGGTGAATGGTGGCCCTCGGAAGCGATTTTGTGCTCTCTCT 291
 Db 122 ---ProValValIleTrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaValPhe 140
 QY 292 GAAGAGTGGGCCCTGGCAGATCATTTGGACTTTTGATGACTATACATCAACCTCTACTCG 351
 Db 141 TyrGluAsnGlyProPheThrIleAlaAsnMetSerLeuValTrpAsnLysPheGly 160
 QY 352 TCGACGAGGTCTCCAAATTTACTATTCCTGTCCTCCAGCCATTGGGAGTCGGCTTTCATAT 411
 Db 467 GlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeuArgArg 484

Db 161 TrpAspLysIleSerAsnIleIlePheValAspGlnProThrGlyThrGlyPheSerTyr 180
 QY 412 AGTGATACGGTGTGAGGTGTCATTAACCTCTACTGGGTGTCGAAAAATTCAGACTTT 471
 Db 181 SerSerAspAspArgAspThrArgHisAspGluThrGlyVal----- 194
 QY 472 GCAGAGTTCAGGCGCGGTACCCCAACCATTTGATGCCACTCTGATCGATACCAATCTT 531
 Db 194 ----- 194
 QY 532 GCCGACAGAGCGGCTGGGAGATCTGCAAGATTCCTTAGTGGACTACCTAGCTTGGAC 591
 Db 195 ---SerAsnAspLeuTyrAspPheLeuGlnValPhePheLysLysHisProGluPhe--- 212
 QY 592 TCTAGGTGTCAGAGTCAAGACTTCAGTCTATGAGGAGGAGAGTATGGAGGAGGAGTATGGT 651
 Db 213 -----IleLysAsnAspPheIleThrGlyGluSerTyrAlaGlyHisTyrIle 229
 QY 652 CTGCAATCTTCAATCATTTTACGACGAGATGAGAGATGCGCAATTAACGGCATCATCGACGCGATC 711
 Db 230 ProAlaPheAlaSerArgValHisGlnGlyAsnLysLys-----AsnGlu 244
 QY 712 GGTGTTACGCTTAATTTCAACTCTCTGGGAATTTAATTAACGGCATCATCGACGCGATC 771
 Db 245 GlyThrHisIleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaIle 264
 QY 772 CAGGCCCTTACTACCTCGTAATTCCTGTGTGAACAATACCTACGGTATCAAGCTCTCAAC 831
 Db 265 GlnTyrLysAlaTyrThrAspTyrAlaLeuGluMetAsnLeuIleGlnLysAspTyr 284
 QY 832 GAGACCGTCTACACTACATGATGAAGTTGCCACCAAAATGCCAAATGGTGGCAGGATTG 891
 Db 285 GluArgIle-----AsnLysPheIleProProCysGluPheAla 297
 QY 892 ATTTCCACTCTGCAACAGACAAACCGC---ACCGCATTAGCTGACTAGCCCTCTGGCC 948
 Db 298 IleLysLeuCysGlyThrAsnGlyLysAlaSerCysMetAlaAlaTyrMetValCysAsn 317
 QY 949 GAAGCCCAACATGTGACGAGGACAATGTGTGAGGGGCCAATACCTACGCCCTTCTGCTGCTG 1008
 Db 318 ThrIlePheAsnSerIleMetLysLeuValGlyThrLysAsnTyr----- 332
 QY 1009 GGTGTGATGATATTCGGCATCCATATGATGATGATGATGATGATGATGATGATGATGATGAT 1065
 Db 333 -----TyrAspValArgLysGluCysGluGlyLysLeuCysTyrAspPheSerAsnLeu 350
 QY 1066 ---AAATTTCTGGCAAGGACTCTGTCTCATGAGCGTATCGCGTC---AACATCACTAC 1119
 Db 351 GluLysPhePheGlyAspLysAlaValArgGlnAlaIleGlyValGlyAspIleGluPhe 370
 QY 1120 ACCGAGTCCAAATGAGACGTCTACTACGCTTTCCAGCAACAGGCGACTTTGTCGCGCC 1179
 Db 371 ValSerCysSerThrSerValTyrGlnAlaMetLeuThrAsp-----TrpMet 386
 QY 1180 AACTTCACTGAA---GACCTCGAGGAGATCTCTGCTCCCGTGGGTGCTCTCCCTCATC 1236
 Db 387 ArgAsnLeuGluValGlyIleProAlaLeuLeuGluAspGlyIleAsnValLeuIleTyr 406
 QY 1237 TATGGCGACCGGATTCATCTCCAACTGGTGTGGCGGTGAGCGCGTTTCCCTCGCTGG 1296
 Db 407 AlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerArgTrpValHisSerMet 426
 QY 1297 ACTACTCCACCGCCCGCTCCGAGGAGGATGAGGAGGATGAGGAGGATGAGGAGGATGAGGAG 1356
 Db 427 GluTyrSerGlyGlnLysAspPheAlaLysThrAlaGluSerSerPheLeuValAspAsp 446
 QY 1357 GTGAGTATGGGAAACTCGGAGTATGGTAAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1416
 Db 447 AlaGlnAlaGlyValLeuLysSerHisGlyAlaLeuSerPheLeuLysValHisAsnAla 466
 QY 1417 GCCCATGAAGTCCCATCTACACCGCCATCGCTCCCTCCCTGCAATTTGTACCGG 1470
 Db 467 GlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeuArgArg 484

```

RESULT 11
ID CBP3_WHEAT
AC CBP3_WHEAT STANDARD; PRT; 500 AA.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y.";
RL J. Biol. Chem. 262:13726-13735(1987).
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- SUBUNIT: Monomer (Probable).
CC -!- INDUCTION: By gibberellic acid (GA).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: J02817; AAA34273.1; -
PIR: A29412; A29412.
HSP: P00729; ICPY.
MEFOPS: S10.009; -.
InterPro: IPR000379; Ser_estrs_site.
InterPro: IPR001563; Serine_carbpept.
Pfam: PF004450; serine_carbpept; 1.
PRINTS: PR00724; CRBOXIPTASEC.
ProDom: PD001189; Serine_carbpept; 1.
PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 21
FT PROPEP 22 73
FT CHAIN 74 484 SERINE CARBOXYPEPTIDASE III.
FT PROPEP 485 500 BY SIMILARITY.
FT ACT_SITE 216 216 BY SIMILARITY.
FT ACT_SITE 404 404 BY SIMILARITY.
FT ACT_SITE 461 461 BY SIMILARITY.
FT BINDING 407 407 SUBSTRATE (BY SIMILARITY).
FT CARBOHYD 144 144 N-LINKED (GLCNAC... ) (POTENTIAL).
FT SEQUENCE 500 AA; B2ACE10EF948ACDA CRG64;
SQ

```

```

Alignment Scores:
Pred. No.:      3-27e-21      Length:      500
Score:          413.50      Matches:     126
Percent Similarity: 41.3%      Conservative: 80
Best Local Similarity: 25.30%      Mismatches:  217
Query Match:      14.24%      Indels:      75
DB:               1          Gaps:          16

US-09-712-338-1_COPY_55_1662 (1-1608) x CBP3_WHEAT (1-500)

Qy      1  CTTCCAGGAAGTACTCCGCGTCCGTGGTAGAGACAGCTACCCAGAACCCACCGGG 60
      ||| ||| |||
Db      47  LeuProGlyArgProArgGlyLeuGly-----AlaGly 58

```

QY	61	GTCAAGACICTTACACCCGCAACAATGTCACCATCCGGTACAGGAACCCGGG---GCA	111
Db	59	AlaGluAspValAlaProGlyGlnLeuLeuGluArgValThrLeuProGlyLeuPro	78
QY	118	GAGGGGTCTGCCAGACTACCCGGGGTGCAAAATCTACTCTGGATAT-----GTCGAC	171
Db	79	GlucLyValGlyAsp-----LeuGlyHisAlaGlyTyrTyrArgLeuPro	94
QY	172	ACCTTCCCGAGTCCCATACCTTCTGCTTCGAAAGCCAGACATAACCCAGAACT	231
Db	95	AsnThrHisAspAlaArgMetPheTyrPhePheGluSerArgGlyLysIysGluAsp	114
QY	232	GCACCTATCACATGTGGTGAAGTGGCCCTGGAGGAGTCTTTGATCGTCTCTTC	291
Db	115	---ProValValIleThrLeuThrGlyProGlyCysSerSerGluLeuAlaValPhe	133
QY	292	GAAGAGTTGGGCCTTGCCATGTCATTCGACTTTTGATGACTACATCAACCCICACTGG	351
Db	134	TyrGluAsnGlyProPheThrIleAlaAsnAsnMetSerLeuValTrpAsnLysPheGly	153
QY	352	TGGAAGAGGTCCTCAAAATTTACTATTCCTGCCAGCCATTCGGAGTGGCGCTTTTCATAT	411
Db	154	TrpAspLysIleSerAsnIleIlePheValAspProAlaThrGlyThrGlyPheSerTyr	173
QY	412	AGTGATACGGTGTAGGGGCCATTAAACCCIGPAACIGGGGTGGTGGAAATTCGAGCTTT	471
Db	174	SerSerAspAspArgAspThrArgHisAspGluAlaGlyVal-----	187
QY	472	GCAGGAGTTACAGGGCGGTACCCAAACCATTGATGCCACTCTGATCGATACTACCAATGTT	531
Db	187	-----	187
QY	532	GCCGAGAGCGCGCTTGGGAGATCCTCGAAGGATTCTTAGTGGACTACCTAGCTTGGAC	591
Db	188	---SerAsnAspLeuTyrAspPheLeuGlnValPhePheLysLysHisProGluPhe---	205
QY	592	TCTAGGGTCAGCTTAAGCACTTCACTATGACGGAGAGCTATGGAGGCACATATGGT	651
Db	206	-----ValLysAsnAspPhePheIleThrGlyGluSerTyrAlaGlyHisTyrIle	222
QY	652	CCTGCATCTCTCAATCATTTTACGACGAGAATGAGAAATGCCAAGCGTAGTGTAAT	711
Db	223	ProAlaPheAlaSerArgValHisGlnGlyAsnLysLys-----AsnGlu	237
QY	712	GGTGTTCAGCTTAATTTCAACTCTCTGGGAATATTAAACGGCATCATCGAGAGCGCATC	771
Db	238	GlyThrHisIleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaIle	257
QY	772	CAGGCCCTTACTACCTCCGTAATTCGTGTGAACAATACCTACGGTATACAGGTGTCAAC	831
Db	258	GlnTyrLysAlaTyrThrAspTyrAlaLeuAsp-----MetAsnLeuIleGln	273
QY	832	GAGACCGTCTACAACTACATCAAGTTTCCCAACCAAAATGCCAAATGGTGTGCCAGATTG	891
Db	274	LysAlaAspLysArgIle-----AsnLysPheIleProProCysGluPheAla	290
QY	892	ATTTCACCTTCCAAACAGACAAACCGC---ACCGCATTTAGTGTACTACGCCCTTCGGCC	948
Db	291	IleLysLeuCysGlyThrAspGlyLysAlaSerCysMetalAlaTyrMetValCysAsn	310
QY	949	GAAACCCACCAATGTCAGGGACAATGTTGAGGGGCCATACTACGCTTGTGTGGTCGT	1008
Db	311	SerIlePheAsnSerIleMetLysLeuValGlyThrLysAsnTyr-----	325
QY	1009	GGTGTGTATGATATTGGCATCCATATGATGACCCGACTCCGCCCAAGTATTACAAC---	1065
Db	326	-----TyrAspValArgLysGluCysGlnGlyLysLeuCysTyrAspPheSerAsnLeu	343
QY	1066	---AAATTTCTGCAAGAGACTTGTCTCATGGACGCTATCGCGGTC---AACATACTAC	1119
Db	344	GluLysPhePheGlyAspLysAlaValArgGluAlaIleGlyValGlyAspIleGluPhe	363
QY	1120	ACCAGTCCAAATATGACGTCTACTAGCTTTCCAGCAACAGGCGACTTGTCTGGCCG	1179

QY	862	AACCAAAATGCCAAATGGTTGC-----CAGGATTTGATTTCCACCTGCAGAAACAGACAAAC	915
Db	217		---
QY	916	CGCACCGCATTTAGCTGACTACGCCCTCTGCGCGGAGCCACCAACATGTCAGGGACAAT	975
Db	231	LyS-----TleLeuAspIleAla---SerThrGluAlaGlyAsnIleAspSerTyrSer	247
QY	976	GTTCAGGGGCCA-----TACTACGCCTTGCTGGT-----	1005
Db	248	IlePheThrProThrCysHisSerSerPheAlaSerSerArgAsnLysValValLysArg	267
QY	1006	---CGTGGTGTGATCATATATTCGGCATCCATATGATGACCGGACTCCGCCAAGT-----	1056
Db	268	LeuArgSerValGlyLysMetGlyGluGlnTyrAspProCysThrGluLysHisSerIle	287
QY	1057	---TATTACACAAATTTCTGSCAAAGGACTCT-----GTCAATGACGCGTATCGCG---	1104
Db	288	ValTyrPheAsnLeuHisGluValGlnLysAlaLeuHisValAsnProValIleGlyLys	307
QY	1105	-----GTCAACATCACTACACCCGACTCCCAATATATGAC	1137
Db	308	SerLysIrrpGluThrCysSerGluValIleAsnThrAsnTrpLysAspCysGluArgSer	327
QY	1138	GTCTACTACGCTTCCAGCAAAACAGCGACTTCTGTCGCCCAACTTCATCGAAGACCTC	1197
Db	328	ValLeuHisIleTyr-----	332
QY	1198	GAGGAGACTCTTGCTCCCGGTGCGTCTCCCTCATCTATGCGCAGCGCGATTACATC	1257
Db	333	HisGluLeuIleGlnTyrGlyLeuArgIleTrpMetPheSerGlyAspThrAspAlaVal	352
QY	1258	TGCAACTGGTTCGGCGGTGAGCGGTTTCCTCGCTGCGAATCTCCAGCGCGCCAG	1317
Db	353	-----IleProValThrSerThrArgTyrSerIleAspAla	364
QY	1318	TTCCGAGCGCAGGTTACAGCCGCCCTCAAGATC-----AACGCGCTCGAGTAT	1365
Db	365	LeuLysLeuProThrValThrProIrrpHisAlaTrpTyrAspAspGlyGluValGly	384
QY	1366	GGGGAACATCCGAGTATGTAATTTCTTCACTCGCTGCTATGAGCGAGCGCATGAA	1425
Db	385	GlyTrpThrGlnGlyTyrLysGlyLeuAsnPheValThrValArgGlyAlaGlyHisGlu	404
QY	1426	GTCCCATATACAGCCCATCGGCTCCCTGCAATGTTTACCGGACTATCTTCGGT	1482
Db	405	ValProLeuHisArgProLysGlnAlaLeuThrLeuIrrpLysSerPheLeuAlaGly	423
RESULT 14			
YSS2_CAEEL	STANDARD;	PRT;	470 AA.
AC	Q09931;		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Putative serine carboxypeptidase K10B2.2 precursor (EC 3.4.16.-)		
GN	K10B2.2.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RA	Miller N.;		
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		

OS Naegleria fowleri.
 OC Eukaryota: Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
 OX NCBI_TaxID=5763;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LEE;
 RX MEDLINE=92267659; PubMed=1587609;
 RT Hu W.-N., Kopachik W., Band R.N.,
 RT "Cloning and characterization of transcripts showing
 RT virulence-related gene expression in Naegleria fowleri.";
 RL Infect. Immun. 60:2418-2424(1992).
 CC -!- FUNCTION: MAY BE REQUIRED BUT IS NOT SUFFICIENT FOR INCREASED
 CC VIRULENCE.
 CC -!- INDUCTION: BY GROWTH ON MAMMALIAN CELLS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M88397; AAA29384.1; -.
 DR PIR; A43828; A43828.
 DR HSP; P10619; LIIV.
 DR MEROPS; S10.UPW; -.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR InterPro; IPR001563; Serine_carbpept.
 DR Pfam; PF00450; serine_carbpept; 1.
 DR PRINTS; PR00724; CRBOXYPTASEC.
 DR ProDom; PD001189; Serine_carbpept; 2.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_S1.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
 KW Hydrolase; Carboxypeptidase.
 FT ACT_SITE 163 163 BY SIMILARITY.
 FT ACT_SITE 399 399 BY SIMILARITY.
 FT ACT_SITE 459 459 BY SIMILARITY.
 SQ SEQUENCE 482 AA; 53848 MW; 0B83049C1A8A0908 CRC64;

Alignment Scores:
 Pred. No.: 7.34e-17 Length: 482
 Score: 351.00 Matches: 139
 Percent Similarity: 35.19% Conservative: 57
 Best Local Similarity: 24.96% Mismatches: 165
 Query Match: 12.09% Indels: 196
 DB: 1 Gaps: 23

US-09-712-338-1_COPY_55_1662 (1-1608) x NF31_NAEFO (1-482)

QY 109 CCGGGGGAGAGGGGGTCTCGGAGACTACCCGGGTGCAAACTCCTACTGGATATGTC 168
 Db 20 ProGlyLeuSerGly-----AsnIleGlyValysSerTyrThrGlyTyrLeu 35
 QY 169 -----GACACCTCCCGAGTCCCATACCTCTCTCGTTCGGAAGCCAGACATAAC 222
 Db 36 LeuAlaAsnAlaThrArgGlyArgTyrLeuPheTyrTrpPheGluSerMetArgAsn 55
 QY 223 CCAGAAACTGCACATATCATCTTGTTGAATGGTGGCCCTCGAAGCGATTCCTTGATC 282
 Db 56 ProSerGlnAspProLeuValMetTrpThrAsnGlyGlyProGlyCysSerSerLeuGly 75
 QY 283 GGTCTCTCGAAGAGTGGGCCCTGCCATCTCAATTCGACTTTTGATGACTACATC--- 339
 Db 76 GlyGluAlaSerGluHisGlyLeuPheLeuValAsnAlaAspGlyAlaThrIleThrArg 95
 QY 340 AACCTCCTCCTCGTGAAGAGGAGGTCTCCAATTTACTATCTGCCCGCAGCCATTGGAGTC 399
 Db 96 AsnProTyrSerIlePasnArgValSerAsnIleLeuTyrIleGluGlnProValGlyVal 115
 QY 400 GGCTTTTCATATAGTATACGGTTTGATGGGTGCCATTAACCTGTAACTGGGTCGTCGAA 459
 |||||||||||||||||||

Db 116 GlyPheSerTyrSerAsnSerThrAsp----- 124
 QY 460 AATTCGAGCTTTGCAGGAGTTTCAGGGCGGTACCCACCATGATGCCACTCTGATCCAT 519
 Db 125 -----Asp 125
 QY 520 ACTACCAATCTT-----GCCGACAGGCGCTTGGGAGATCCTGCAAGGATTC 567
 Db 126 TyrGlnAsnLeuAsnAspValGlnAlaAlaSerAspMetAsnAsnAlaLeuArgAspPhe 145
 QY 568 CTTAGTGGACATACCTAGCTTGGACTCTAGGTCAGTCTAAGGACTTCAAGTCTAGTCTAGTACG 627
 Db 146 LeuThrArgPheProGlnPheIleGlyArg-----GluThrTyrLeuAlaGly 161
 QY 628 GAGAGCTATPGAGGCGACTATGGTCT---GCATCTTCAATCAATTTTACGACGAGAT 684
 Db 162 GluSerTyrGlyValTyrValProThrThrAlaTyrAsn----- 175
 QY 695 GAGAGAAATGCCAAGGTAGTGTATATGGTGTACCTTAATTTCAACTCTCTGGCAAT 744
 Db 176 -----IleValIleGlyAsnGlyLysGlyGlnGlnProTyrValAsnLeuValGlyIle 193
 QY 745 AAT-----AACGGCATCATCGAC----- 762
 Db 194 LeuValGlyAsnGlyValThrAspAlaGluAlaAspSerAsnSerIleProPrometMet 213
 QY 763 -----GAGGCGATCCAGGCCCTTACTACCTGGAATTCGCTGTGAACATACCTAC 813
 Db 214 LysTyrHisSerLeuIleSerIleLysIleTyrGluGlu----- 226
 QY 814 GGTATCAAGGTGTCAACGAGACCGTCTACAACATGATAGTTCGCAACCAATGCCA 873
 Db 227 GlyTyrLysAlaCysGlnGlyAspPheTyr-----AlaAsnGlnAsnLeu 241
 QY 874 AATGGTTCGACGAGTTCATTCACCTGCAACAGACAAACCGCCGATTCAGTGCAC 933
 Db 242 ProAlaCysGlnLysPheLeuThr----- 249
 QY 934 TAGCCCTCTCGCCGAGCCACCAACATGTGACGAGGACAATTTGAGGGGCCATCTACTAC 993
 Db 250 -----AspSerSerAsn---AlaMetGlyAsnIle---AsnProTyrTyr 262
 QY 994 GCCTTT----- 999
 Db 263 IleTyrAspSerCysProTyrLeuGlyIleAsnLeuGlnGlnLysLeuLysThrThrGln 282
 QY 999 ----- 999
 Db 283 GluMetThrPheGlnValLeuAspProLysThrGlnGlnProValLysIleHisProLeu 302
 QY 1000 -----GCTGGTCTGGTGTATGAT 1020
 Db 303 PheGlnMetTyrLysHisGlyTyrSerLysArgValAlaAsnGluArgAsnPheAla 322
 QY 1021 ATTTCGGCATCCATATGATGACCGACCTCCGCAAGTTAT---TACAACAATTTCTGGCA 1077
 Db 323 ProArgPheGluThrAspAlaProCysValProAsnGlnSerIleAlaLysTyrPheArg 342
 QY 1078 AAGGACTCTGTGAGGACGCTATCGGCTCAAC----- 1110
 Db 343 ArgLeuAspValGlnGlnAlaLeuGlyValArgArgLysThrAlaAspProAsnGlyTyr 362
 QY 1111 -----ATCACTACACCCAGTCGAATATGACGTC-----TACTAC 1146
 Db 363 AsnIleCysThrGlyIleIleAsnTyrThrGlnValTyrSerThrIleLeuProPheTyr 382
 QY 1147 GCTTTCAGCAAAACAGGCGACTTGTCTGGCCCACTTCATCGAAGACCTCGAGGAGATC 1206
 Db 383 Ala----- 383
 QY 1207 CTTGCTCTCCCC---GTCCGTGTCTCCCTCATCTATGCGCGACGCCGATTCATCTGCAAC 1263
 Db 384 LysLeuLeuProHisIleArgIleLeuValTyrSerGlyAspThrAspMetValValAsn 403
 |||||||

```
Oy 1264 TGGTTCGGCGGTACAGCGGTTTCCTCGCTCGAACTACTCCCAAGCGCCAGTCCGA 1323
    ||| |||||
Db 404 GlyLeuGlyThrGlnAlaAlaIleAspLysLeuGlnLeuGlnGluThrSerSerIrpArg 423
    : : : : :
Oy 1324 AGCGCAGGTACACGCCCTGAAAGTCAACGGCGTCGAGTATGGGAAACTCGCGAGTAT 1383
    : : : : :
Db 424 ThrTrpGluPheAspSer---AlaLeuGlyThrValValGlyGlyTyrIleArgLysPhe 442
    : : : : :
Oy 1384 -----GGTAATTCTCTCTCACTCGCGTCTATGAGCGAGCCCATGAAGTCCCA 1431
    : : : : :
Db 443 GluLysSerGlyLysGlyLeuThrPheIleThrValArgGlyAlaGlyHisMetValPro 462
    : : : : :
Oy 1432 TACTACAGCCCATCGGCTCCCTGCAATTGTTTAACCGGACTATCTCGGT 1482
    : : : : :
Db 463 LeuValLysProAspSerAlaPheTyrMetPheLysAsnPheIleAspGly 479
    : : : : :
```

Search completed: September 16, 2003, 18:43:07
Job time : 58 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:14:18 ; Search time 111.5 Seconds
(without alignments)
7443.024 Million cell updates/sec

Title: US-09-712-338-1_copy_55_1662

Perfect score: 2904

Sequence: 1 cttccagggaagtacacggc.....gcattgtccagtgttggtatg 1508

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=framet_n2p.model -DRV=xlp
-Q=/cgn2.1/USPTO.spool.p/US09712338/runat_16092003_144325_14356/app_query.fasta_1.1799
-DB=SPREMBL_23 -QFWT=fastan -SURF=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338.ecgn_1.123.0runat_16092003_144325_14356 -NCPD=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPREMBL_23:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_virus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2875	99.0	555	3 Q96VZ9	Q96vz9 aspergillus

2	746.5	25.7	460	3 Q12569	Q12569 absidia zyc
3	601	20.7	552	3 Q96VC4	Q96vc4 emericella
4	596.5	20.5	541	3 Q94152	Q94152 pichia angu
5	595.5	20.5	537	3 Q14414	Q14414 pichia angu
6	479	16.5	507	10 Q8VWQ0	Q8vwq0 gossypium h
7	473	16.3	429	10 Q8GVU1	Q8gvt1 oryza sativ
8	466.5	16.1	505	10 Q9FFB0	Q9ffb0 arabidopsis
9	461	15.9	501	10 Q9XH61	Q9xh61 matricaria
10	460	15.8	508	10 Q8L6A7	Q8l6a7 theobroma c
11	439	15.1	510	10 Q9LXH4	Q9lxh4 arabidopsis
12	438	15.1	510	10 Q932C3	Q932c3 arabidopsis
13	413.5	14.2	2105	5 Q17679	Q17679 caenorhabdi
14	406.5	14.0	499	10 Q9FMX9	Q9fmx9 arabidopsis
15	385	13.3	490	10 Q9FYF7	Q9fyf7 oryza sativ
16	383	13.2	482	10 Q9SV04	Q9sv04 arabidopsis
17	376.5	13.0	524	10 Q8G1K2	Q8gtk2 oryza sativ
18	363	12.5	480	10 Q9LEV1	Q9lev1 arabidopsis
19	362.5	12.5	487	10 Q9SV03	Q9sv03 arabidopsis
20	360	12.4	437	10 Q9FWG1	Q9fwg1 oryza sativ
21	360	12.4	482	10 Q9FRJ0	Q9frj0 oryza sativ
22	358	12.3	470	10 Q22732	Q22732 arabidopsis
23	357.5	12.3	479	10 Q949Q7	Q949q7 arabidopsis
24	355.5	12.2	360	10 Q9M450	Q9m450 cicer ariet
25	349	12.0	474	11 Q9D2D1	Q9d2d1 mus musculu
26	348	12.0	452	10 Q9ZQ00	Q9zq00 arabidopsis
27	344	11.8	465	10 Q04084	Q04084 arabidopsis
28	344	11.8	494	10 Q9FEU4	Q9feu4 pisum sativ
29	344	11.8	501	10 Q9SV02	Q9sv02 arabidopsis
30	342.5	11.8	471	10 Q8L9Y0	Q8l9y0 arabidopsis
31	342.5	11.8	473	10 Q9SCA9	Q9scs9 arabidopsis
32	342.5	11.8	669	10 Q8LPY6	Q8lpy6 oryza sativ
33	340	11.7	512	5 Q76725	Q76725 caenorhabdi
34	337.5	11.6	425	10 Q65568	Q65568 arabidopsis
35	333.5	11.5	510	3 Q60123	Q60123 schizosacch
36	333	11.5	487	10 Q9MUG3	Q9mug3 arabidopsis
37	332	11.4	465	10 Q9M099	Q9m099 arabidopsis
38	331	11.4	497	10 Q8L7B2	Q8l7b2 arabidopsis
39	330.5	11.4	472	10 Q9LSM9	Q9lsm9 arabidopsis
40	330.5	11.4	474	10 Q82228	Q82228 arabidopsis
41	330	11.4	445	5 Q814E3	Q814e3 caenorhabdi
42	330	11.4	1203	5 Q45916	Q45916 caenorhabdi
43	329.5	11.3	548	4 Q9BR08	Q9br08 homo sapien
44	328.5	11.3	434	11 Q9D3S9	Q9d3s9 mus musculu
45	328	11.3	504	10 Q9LSV8	Q9lsv8 arabidopsis

ALIGNMENTS

RESULT 1
Q96VZ9
ID Q96VZ9 PRELIMINARY; PRT; 555 AA.
AC Q96VZ9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase S1 precursor (EC 3.4.16.6).
GN CPl.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK3;
RA van den Broek P.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF394242; AK77166.1; -
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER; 1.
KW Carboxypeptidase; Hydrolase; Signal.

FT	SIGNAL	1	18	POTENTIAL.	
SQ	SEQUENCE	555	AA;	61168	MW; 456B63B0CEB55222 CRC64;
Alignment Scores:					
Pred. No.:	5.85e-198	Length:	555		
Score:	2875.00	Matches:	535		
Percent Similarity:	100.00%	Conservative:	1		
Best Local Similarity:	99.81%	Mismatches:	0		
Query Match:	99.00%	Indels:	0		
DB:	3	Gaps:	0		
US-09-712-338-1_copy_55_1662 (1-1608) x Q96V29 (1-555)					
QY	1	CHTCCAGGAGTACACCGGCGCTCGCTGCTAGAACAGCTACCCCAAGAACCCACCGGG	60		
DB	19	LeuProGlySerThrProAlaSerValGlyIleArgGlnLeuProLysAsnProThrGly	38		
QY	61	GTCAGACACTTACAAACCGCAACATGTCACATCCGGGTACAAAGAACCCGGGGCAGAG	120		
DB	39	ValLysThrLeuThrThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu	58		
QY	121	GGCGTCTGGAGACTACCGGGGTGCAAAATCTACTCTGGATATGTCACACCTCTCCC	180		
DB	59	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro	78		
QY	181	GAGTCCCATCACTTCTCTGTTCTTCGAAAGCCAGACATAACCCAGAAACTGCACCTATC	240		
DB	79	GluSerHisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle	98		
QY	241	ACATTGTGGTGAATGTGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGGAAGATTG	300		
DB	99	ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeu	118		
QY	301	GGCCCTTGCCATGCAATTCGACTTTTGATGACTACATCAACCTCACTCGTGAAGCAG	360		
DB	119	GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu	138		
QY	361	GTCTCCAAATTACTATCTCTGCCAGCCATGGGAGTCGGCTTTTCATATAGTATACG	420		
DB	139	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrAsnAspThr	158		
QY	421	GTTGATGGGTCATTAACCCCTGTAACGTGGGTGCTCGAAATTCGAGCTTTGAGGAGTT	480		
DB	159	ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal	178		
QY	481	CAGGGCCGGTACCAACCATGTATGCCACTCTGATCGATACCAATCTTGGCCGACAG	540		
DB	179	GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGlu	198		
QY	541	GCCGCTTGGGAGATCTCGAAGGATCTTWTAGTGGACTACCTAGCTTGGACTCTAGGGTG	600		
DB	199	AlaAlaThrGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal	218		
QY	601	CAGCTCAAGGACTTCAGTCTATGACGACGAGAGCTATGAGGGGCATATGTCCTGCATTC	660		
DB	219	GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe	238		
QY	661	TTCAATCATTTTACGACCAAGTACAGAGATTCGCAACGGTAGTGTATATGTTGTTCCAG	720		
DB	239	PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln	258		
QY	721	CTTAATTTCAACTCTCTGGGAATTATTAAACGGCATCAFCGACGAGCGGATCCAGGCCCT	780		
DB	259	LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro	278		
QY	781	TACTACCTGAATTCGCTGTGACAAATACCTACGGTATCAGGGCTGCTCAACGACGCGTC	840		
DB	279	TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal	298		
QY	841	TACAACTACATCAAGTTTGCAACCAATGCCAATGTTGCCAGGATTTGATTTCCACC	900		
DB	299	TyrAsnTyrMeClyPheAlaAsnGlnMetProAsnGlyCysGluAspLeuIleSerThr	318		

QY	901	TGCAACAGACAGAAACCGCACCGCATTAGCTAGTACGCCCTCTGCGCGAAGCCACCAAC	960
DB	319	CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn	338
QY	961	ATGTGCAGGACAAATGTTGAGGGCCATACCTAGCGCTTTGCTGGTCTGGTGTATGAT	1020
DB	339	MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp	358
QY	1021	ATTGGCATCCATATGATGACCGGACTCCGCCCAAGTATTACAAACAAATTTCTGGCAAAG	1080
DB	359	IleArgHisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLys	378
QY	1081	GACTCTGTCATGACGCTATFCGCGCTCAACATCACTACACCCAGCTCCCAATATGAGCTC	1140
DB	379	AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspVal	398
QY	1141	TACTAGCTTTCCAGCAAAACAGCGCTTTGCTGCGCCCAACTTCATCGAAGACCTCGAG	1200
DB	399	TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu	418
QY	1201	GAGATCTTGTCTCTCCCGCTGCGTCTCTCCCTCATCTATGCGGACGCGGATTCATCTGC	1260
DB	419	GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys	438
QY	1261	AACTGGTTCGGCGCTCAGGCGCTTCCCTCGCTCGGAACCTACTCCCAAGCGCCGACGTC	1320
DB	439	AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPhe	458
QY	1321	CGAAGCGCAGGGTACACGCGCCCTGAAAGTCAACGCGCTCGAGTATGGGAAACATCGCGAG	1380
DB	459	ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu	478
QY	1381	TATGGTAATTTCTCTTCACTCGCGTCTATGAGCGCGCCATGAAGTCCCATACACAG	1440
DB	479	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	498
QY	1441	CCCATCCCTCCCTCGCAATTTTAAACGGGACTATCTTCGGTTGGGATATCGCAGAGGC	1500
DB	499	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly	518
QY	1501	CAGAAGAGATCTGCGCCAGCTACAGACGAAATGAAGCGGTACAGCTACGCTACGATACAG	1560
DB	519	GlnLysLysIleTyrProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln	538
QY	1561	TCGTCGCGTCCGCTGCGCTACGCGCTACGAGCATGCCAGTGTCCAGTGTGTATG	1608
DB	539	SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet	554

RESULT 2

ID	Q12569	PRELIMINARY;	PRT;	460	AA.
AC	Q12569;				
DT	01-NOV-1996	(TReMBLrel. 01, Created)			
DT	01-NOV-1996	(TReMBLrel. 01, Last sequence update)			
DT	01-JUN-2002	(TReMBLrel. 21, Last annotation update)			
DE	Prepro-carboxypeptidase 2 (EC 3.4.16.5).				
GN	spcz.				
OS	Absidia zychoae.				
OC	Eukaryota; Fungi; Zygomycota; zygomycetes; Mucorales; Mucoraceae;				
OC	Absidia.				
OX	NCBI_TaxID=36079;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NRIC 1199;				
RC	MEDLINE=95308538; PubMed=7788719;				
RA	Lee B.R., Takeuchi M., Kobayashi Y.;				
RI	"Molecular cloning and sequence analysis of the sopz gene encoding the				
RL	serine carboxypeptidase of Absidia zychoae.";				
DR	Curr. Genet. 27:159-165(1995).				
DR	EMBL; D16519; BAA03966.1; -.				
DR	HSSP; P00729; 1YSC.				
DR	MEROPS; S10.010; -.				
DR	InterPro; IPR001563; Serine_carbpept.				

InterPro: IPR000379; Ser_estrs_site.
 PRAM; PF00450; serine_carbpept; 1.
 DR Pfam; PR00724; CRBOXYPTASEC.
 DR ProDom; PD001189; Serine_carbpept; 1.
 KW Carboxypeptidase; Hydrolase.
 FT CHAIN 52 460 MATURE CARBOXYPEPTIDASE Z.
 SQ SEQUENCE 460 AA; 50825 MW; CE3A410764548BB6 CRC64;

Alignment Scores:
 Pred. No.: 3-65e-45 Length: 460
 Score: 746.50 Matches: 175
 Percent Similarity: 50.31% Conservatives: 71
 Best Local Similarity: 35.79% Mismatches: 182
 Query Match: 25.71% Indels: 61
 DB: 3 Gaps: 14

US-09-712-338-1_COPY_55_1662 (1-1608) x Q12569 (1-460)

QY	16	CGCGCGTCGTCGGTAGAGACAGACTACCCAGAAACCCACCGGGTCAAGACTCTTACA	70	...
Db	29	ProMetSerHisValLeuArgGlnAspGlyAsnAspThr
QY	76	ACCGCAACAAFTGCCATCCGCTACAGAAACCCGGGCGAGGCGCTCGCGAGACT	135	...
Db	44	SerGlyAsnThrThrGlnLeuLysTyrThrSerPro
QY	136	ACCCCGGGTGCAAACTCTACTCTGGATATCTGCACACCTCTCCGAGTCCCATACCTC	195	...
Db	60	---ProAspValLysGlnTyrSerGlyTyrLeuAspAlaAlaAsnAspGluHisTyrPhe	78	...
QY	196	TTCCTGGTCTTCGAAGCCAGACATACCCAGAACTGCATCATCATTTGTGGTTC	255	...
Db	79	PheTyrPheGluSerLysAsnAspProLysAsnAspProLeuThrIleTyrLeuAsn	98	...
QY	256	GTGCGCCCTGGAAGGATCTTTGATCGGCTCTCTCCAGAGTTGGCCCTTCGCATGTC	315	...
Db	99	GlyGlyProGlyCysSerSerLeuIleGlyTyrGluLeuGlyProCysGlnGln	118	...
QY	316	AATCGACITTTTGATGATACATCAACCCCTCACTCGTGGAAACAGAGGCTCTCCAATTTACTA	375	...
Db	119	AsnGlySer
QY	376	TTCCTGTCGCCACCATGGGAGTCGGCTTTTCATATAGTGTATAGTGGTCCAT	435	...
Db	135	PhePheAspGlnProAspGlyValGlyPheSerTyr
QY	436	AACCCTGTAACCTGGGGTCGTGAAATTCGAGCTTCGAGGAGGTCAGGCCGCGTACCCA	495	...
Db	147	---
QY	496	ACCAITGATGCCACTCTGTATGATACCTACCAANTCTTCCGCGAGGCGCGTGGGAGATC	555	...
Db	150	---
QY	556	CTGCAAGGATTCCTTAGTGACTACTAGTGTGGACTCTAGGTCAGAGCTCAAGGACTTC	615	...
Db	165	LeuGlnAlaPheTyrGluThrPhePro
QY	616	AGTCTATGCGGAGAGCATGATGAGGCGCATATGGTCCCTGCATCTTCATCATTTTAC	675	...
Db	181	HisTyrPheGlyGluSerTyrGlyGlyHisTyrIleProGlyPheAlaSerHisValVal	200	...
QY	676	GAGCAGAAATGAGAAATTGCCACGGTAGTGTAAATGGTGTTCAGCTTAATTCACCTCT	735	...
Db	201	AspMetAsnLysLysValGlnSerGlyGluGluLysValValProLeuLysSer	220	...
QY	736	CTGGCAATTATTAACGGCATCATCGAGGCGGATCCAGGCCCTTACTACCTCGAATTC	795	...
Db	221	IleGlyValGlyAsnGlyPheIleAspAlaValIleGlnTyrLysSerTyrProLysMet	240	...
QY	796	GCTGTGAACATACTACGGTATCAGGCGTGTCACGAGACCGCTCTACAACTACATGAG	855	...
Db	241	ThrCysHisSerThrTyrProAlaValLeuSerGluGlu

Alignment Scores:

Pred. No.: 2,27e-34 Length: 541
 Score: 596.50 Matches: 164
 Percent Similarity: 46.14% Conservative: 63
 Best Local Similarity: 33.33% Mismatches: 184
 Query Match: 20.54% Indels: 81
 DB: 3 Gaps: 20

US-09-712-338-1_copy_55_1662 (1-1608) x 094152 (1-541)

```

QY 82 AACATGTCACATCCGCTACAG-----CAACCCGGGAGAGGGCTGTGGAGACT 135
DB 115 SerAsnTyrLysLeuArgValLysLysThrProGlyAlaLeuGluLeuAspHisThr 134
QY 136 ACCCGGGGTGCAATCTACTCTGGATATCTGACACCTCTCCCGAG---TCCCATACC 192
DB 135 -----ArgGlnTyrSerGlyTyrLeuAspValGluAspGluAspLysHisPhe 150
QY 193 TTCCTCTGGTCTTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTTGTGGTG 252
DB 151 PheTyrIrpMetPheGluSerArgAsnAspProAlaAsnAspProValIleLeuTyrLeu 170
QY 253 AATGCTGGCCCTGGAAGCGATTCTTGTAGTGGTCTCTTCGAAGAGTGGGCCCTTGGCAT 312
DB 171 AsnGlyGlyProGlyCysSerSerLeuThrGlyMetLeuPheGluLeuGlySerAlaSer 190
QY 313 GTCAATTTCGACTTTTGTGACTACATCAACCTCCTCTGTCGAGAGGCTCCAAATTA 372
DB 191 IleGlyProAspLeuLysProIleAsnAsnProTyrSerIrpAsnSerAsnAlaThrVal 210
QY 373 CTATCTCTGCCAGCCATTGGGAGTCGGCTTTTCATATAGTATGATGACGGTGTGGTGC 432
DB 211 IlePheLeuAspGlnProValAsnValGlyPheSerTyrSer----- 224
QY 433 ATTAACCCCTGAATCGGGCTGTCGAAATTCGAGCTTCGAGAGTTCAGGGCCGGTAC 492
DB 224 ----- 224
QY 493 CCAACCATTTGATGCGACTCTGATGATACCAATCTTGGCGAGAGCGGCTTGGAG 552
DB 225 -----SerLysSerValSerAsnThrValAlaAlaGlyLysAspValTyrAla 240
QY 553 ATCTGCAAGGATTCCTTACTGGACTACCTAGCTTGGACTAGAGTGCAGTCTAAGCAG 612
DB 241 PheLeuGluLeuPheTyrGlnGlnPheProHisLeu-----LeuLysAsnAsp 256
QY 613 TTCAGTCTATGAGGAGAGCTATGGAGGCACTATGCTCCTGCATCTTCATCATATTT 672
DB 257 PheHisIleAlaGlyLeuSerTyrGlyGlyHisTyrIleProValPheAlaSerGluIle 276
QY 673 TAGCAGCAGAAATGAGAGAAATGCCAAGGTAGTGTAAATGGTGTTCAGCTTAATTTCAAC 732
DB 277 LeuThrHisAlaAspArg----- 287
QY 733 TCCTCTGGATATTAACGGCATCATCGAGAGCGGATCCAGGCCCTTACTACCTCGAA 792
DB 288 SerValLeuIleGlyAsnGlyLeuThrAspProLeuAsnGlnTyrProPheTyrGluArg 307
QY 793 TTCGCTGTGAACAATACCTACGGTATCAAG---CGTGTCAACGAGACCGTCTCACTAC 849
DB 308 MetAlaCysSerThrAspGlyTyrGluProThrLeuAspGluSerGluCysGluGly 327
QY 850 ATGAAGTTTGCACCAAAATGCCAAATGGTTGCGAGAGATTGTATTTCCACTCCAAACAG 909
DB 328 Met-----LeuGluThrLeuProArg---CysLeuSerLeuIleGluSerCysTyrSer 344
QY 910 ACAACCCGACCCGATTAAGTACTACGCCCTCTGGCCGGAAGCCACCAACATGTGCAGG 969
DB 345 SerGln-----SerValPheSerCysValProAlaSerIleTyrCysAsn 359
QY 970 GACAATGTTGAGGGCCATACAGCTTGTCTGGTGTGGTGTGTATGATATTCGGCAT 1029
DB 360 AsnAlaGlnLeuGlyProPheGlnLys---ThrGlyArgAsnValTyrAspValArgLys 378

```

```

QY 1030 CCA-----TATGATGACCCGACTCCGCCAAGTTATATACACAATTT 1071
DB 379 MetCysGluGlyThrLeuCysTyrLysAsp-----MetGluTyrIleAspGlnTyr 395
QY 1072 CTGGCAAGGACTCTGTATGACGCTATCGCGGTCAACATC---AATACACCCGAGTCC 1128
DB 396 LeuAsnGlnAspPheValLysGluLysIleGlyAlaGluValGluThrTyrGluSerCys 415
QY 1129 AATAATGACGTCTACTACGCTTCCAGCAAAACAGCGACTTGTCTGGGCC----- 1179
DB 416 AsnPheAspValAsnArgAsnPheLeuPheAlaGlyAspIrpMetLysProTyrHisLys 435
QY 1180 AACITTCATCGAAGACTCGAGGAGATCTCTCTCCCGCGTCTCTCCCTCATCAT 1239
DB 436 AsnValIleAsnLeuLeuGluGln-----GlyLeuProVal-----LeuIleTyr 450
QY 1240 ---GGCGAGCGGATTACTGCAACTGGTTCGCGCGGTGAGCGGTTTCCCTCCTCGCG 1296
DB 451 AlaGlyAspLysAspPheIleCysAsnIrpLeuGlyAsnGlnAlaIrpSerAsnGluLeu 470
QY 1297 AACTACTCCCAAGCGCGCTCGAAGCGAGCGG---TACACGCCCTGAAAGTCAAC 1353
DB 471 ProIrpSerGlyHisGluGluPheGluSerAlaGluLeuTyrAsnLeuThrLeuLysAsp 490
QY 1354 GCGCTCGAGTATGGGAACTCGCGAGTATGTAATTTCTCTTCTACCTCGGCTCATGAG 1413
DB 491 GlyThrLysValGlyGluValLysAsnAlaGlyLysPheThrPheAlaArgMetPheAsp 510
QY 1414 GCAGGCCAAGAATCCCATACTACAGCCCATCGCTCTCCCTCCCAATTTTAAACGGGACT 1473
DB 511 GlyGlyHisMetValProTyrAspGlnProGluSerSerLeuAlaMetValAsnArgIrp 530
QY 1474 ATCTTCGTTGGGATATCGCAGAGGCGCAGAGAAG 1509
DB 531 IleAlaGly---AspTyrSerLeuGlyThrLysLys 541

```

RESULT 5

```

ID O14414 PRELIMINARY; PRT; 537 AA.
AC O14414;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DI 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase Y.
GN CPY.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4905;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DL-1;
RA Bae J.-H., Kim H.-Y., Sohn J.-H., Choi E.-S., Rhee S.-K.;
RT "Molecular cloning and characterization of a gene coding for
RI carboxypeptidase Y from yeast Hansenula polymorpha DL-1.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U67174; AAB68520.2; -.
DR HSSP; P00729; ICPI.
DR MEROPS; S10.001; -.
DR InterPro; IPR001563; Serine_carbpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 537 AA; 60270 MW; FC0LE57D7582C941 CRC64;

```

Alignment Scores:
 Pred. No.: 2.67e-34 Length: 537
 Score: 595.50 Matches: 165
 Percent Similarity: 45.01% Conservative: 65
 Best Local Similarity: 32.29% Mismatches: 196

Query Match:	20.51%	Indels:	85
DB:	3	Gaps:	20
US-09-712-338-1_COPY_55_1662 (1-1608) x 014414 (1-537)			
QY	37	CAGCTACCAAGAACCCACCCGGGTCAGAGCTCTTACAAACCGCAACAATGTCACCATC	96
DB	92	LysilethrlyslsProAlaAspPheTrpAspPheAsnValLeuAsnGluLysPheSer	111
QY	97	CGGTACCAAG-----GACCCGGGCGAGGCGGCTCGAGACATCC	138
DB	112	AsnTyrLysLeuArgValLysLysThrAspProGlyAlaLeuGlyLeuAspHisThr---	130
QY	139	CCGGTGTCAAACTCTGATATGTCAGACACCTCTCCCGAG---TCCCATACCTTC	195
DB	131	-----ArgGlnIyrSerGlyTyrLeuAspValGluAspGluAspLysHisPhePhe	147
QY	196	TTCTGTTCTCCAGACGACATACCCAGAACTGCACCTATACATGTTGTTGTAAT	255
DB	148	TyrTrpMetPheGluSerArgAsnAspProValAsnAspProValLleLeuTrpLeuAsn	167
QY	256	GTTGGCCCTGAGACCGATTCTTTCATCGGTCTCTTCAGAGAGTTGGCCCTTGCCATGTC	315
DB	168	GlyGlyProGlyCysSerSerLeuThrGlyMetLeuPheGluLeuGlySerAlaSerIle	187
QY	316	AATTCCGACTTTGATGACTACATCAACCCCTCACCTCGTGGAAACGAGGCTCCAAATTACTA	375
DB	188	GlyProAspLeuLysProIleAsnAsnProTyrSerTrpAsnSerAlaThrValle	207
QY	376	TTCTGTCCGACGATTTGGAGTCGGCTTTTCATATAGTAGTACAGGTTGATGGTCCATT	435
DB	208	PheLeuAspGlnProValAsnValGlyPheSerTyrSer-----	220
QY	436	AACCTGTAACTGGGTGCTGCAAAATTCGAGCTTTGCAGGAGTTCAGGGCGGTCACCA	495
DB	220	-----	220
QY	496	ACCATTTGATGCACTCTGATGATACTACCAATCTTCCGAGAGCGCGCTTGGAGATC	555
DB	221	-----SerLysSerValSerAsnThrValAlaAlaGlyLysAspValTyrAlaPhe	237
QY	556	CTGAAGATTCCTTAGTGGACTACCTAGCTGTTGGAGCTAGGGGCGAGCTCAAGAGCTC	615
DB	238	LeuGluLeuPheTyrGlnGlnPheProHisLeu-----LeuLysAsnAspPhe	253
QY	616	AGTCTATGGACGAGCTATGGAGGCGACTATGCTCTGCACTTCTCAATCATTTTAC	675
DB	254	HisIleAlaGlyGluSerTyrGlyGlyHisTyrIleProValPheAlaSerGluLeuLeu	273
QY	676	GAGCAGAATGAGAGAAATGCCAACGGTAGTGTGTTAATGGTGTCTCAATTTCAACTCT	735
DB	274	ThrHisAlaAspArg-----SerPheAsnLeuThrSer	284
QY	736	CTGGGAATTAACGGCATCATCAGAGGGATCCAGGCCCTTACTACCTGAATTC	795
DB	285	ValLeuIleGlyAsnGlyLeuThrAspProLeuAsnGlnTyrProPheTyrGluArgMet	304
QY	796	GCTGTGACAAATACCTACGAGTATCAAG--GCTGTCAACGAGACCGTCTCAACATACATG	852
DB	305	AlaCysSerThrAspGlyGlyTyrGluProThrIleAspGluSerGluCysGluGlyMet	324
QY	853	AAGTTTGCAACCAATGCCAAATGGTTGCCAGGATTGATTCACCTGCCAGCAACAGACA	912
DB	325	-----LeuGluThrLeuProArg---CysLeuSerLeuIleGluSerCysTyrSerSer	341
QY	913	AACCCACCGCATTTAGCTACGATACCCCTCTCCGCGGAGCCACCACCAATGTCAGGGAC	972
DB	342	Gln-----SerValPheSerCysValProAlaSerIleTyrCysAsnAsn	356
QY	973	AATGTTGAGGGGCCATACATACGCCCTTGTGCTGCTGGTGTGTATGATATTCGGCATCCA	1032
DB	357	AlaGlnLeuGlyProPheGlnLys---ThrGlyArgAsnValTyrAspValArgLysMet	375

Alignment Scores:
Pred. No.: 6,06e-26 Length: 507
Score: 479.00 Matches: 145

Percent Similarity: 44.27% Conservative: 83
 Best Local Similarity: 28.16% Mismatches: 191
 Query Match: 16.49% Indels: 96
 DB: 10 Gaps: 20

US-09-712-338-1_COPY_55_1662 (1-1608) x Q8VW00 (1-507)

```

QY 31 AGAAGACAGCTACCCAGAACCCACCGGGGTCAAGACTCTTACAAACCGCAACAATGTC 90
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   ArgValSerLeuProGluGlyProLysLeuValGluLys----- 77

QY 91 ACCATCCGGTACAGAGAACCGGGGCGAGAGGGCGCTCCGAGACTACCCGGGGTGCAAA 150
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   -----ArgPheLysPheProThrLeuGlu-----ValProGluGlyVal 90

QY 151 TCCTAC-----TCTGGATAT-----GTCGACACTCTCCCGAGTCC 186
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   91 SerPheGluAspLeuGlyHisHisAlaGlyTyrTyrLysLeuProAsnSerHisAspAla 110

QY 187 CATACCTCTCTCGGTTCTTCAAGCCAGACATACCCAGAACTGCACCTATCACATTG 246
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   111 ArgMetPheTyrPhePheGluSerArg---AsnSerLysLysAspProValValle 129

QY 247 TGTGTGAATGGTGGCCCTGAAGGATCTTGTGATCGGTCCTCTTCGAGAGATTTGGCCCT 306
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   130 TrpLeuThrGlyGlyProGlyCysSerSerGluLeuAlaLeuPheTyrGluAsnGlypro 149

QY 307 TGGCATGCTCAATTCGACTTTTGTGATGACTACATCAACCTCCTCTCGTGGACGAGTCTCC 366
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   150 PheThrIleAlaAspAsnMetSerLeuValTrpAsnGluTyrGlyTrpAspLysAlaSer 169

QY 367 AATTACTATTCCTGCTCCAGCATTTGGAGTGGGCTTTTCATATACATGATACGGTGTAT 426
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   170 AsnLeuLeuTyrValAspGlnProGlyGlyThrGlyPheSerTyrSer----- 185

QY 427 GGTCCATTACCTGTACTGGGTCGTCGAAATTCGAGCTTTCGAGAGTTCAGAGGC 486
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   186 -----SerAspGlnArgAspIleArgHis 193

QY 487 CGGTACCCCAACCATTCGACCTGATCGATACCTACCAATCTTCGCCGAGAGCGGCT 546
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   194 AsnGluAspGluValSerAsnAspLeuTyrAsp----- 204

QY 547 TGGGAGATCGTCAAGATTCCTTAGTGGACTACCTAGTGTGGACTCTAGGTGTCAGTCT 606
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   205 -----PheLeuGlnAlaPheAlaGluHisProGluPhe-----AlaLys 218

QY 607 AAGGACTTCAGTCTATGGACGAGACTATGGAGGCACTATGGTCTCGATTTCTTCAAT 666
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   219 AsnAspPhePheIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaPheAlaAla 238

QY 667 CATTTTTCAGCAGCAATCAGAGAAATGCAACGGTAGTGTAAATGGTCTCAGCTTAAT 726
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   239 ArgValHisArgGlyAsnLys-----AlaLysGluGlyIleHisIleAsn 253

QY 727 TTCACCTCTCTGGAAATTTAAACGGCAATCATCGAGAGGCGATCCAGGCGGCTTACTAC 786
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   254 LeuLysGlyPheAlaIleGlyAsnGlyLeuThrAspProAlaIleGlnTyrLysAlaTyr 273

QY 787 CTTGAATTCGCTGACACATACCTACCGTATCAGGCTGTCAACGAGACCGTCTACAC 846
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   274 ThrAspTyrAlaLeuAspMetGlyValIleLysLysSerAspHisAspArgIle----- 291

QY 847 TACATGAGTTTGGCAACCAATGCCAAATGGTTCGACGAGATTGATTTCCACCTGCAAA 906
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   292 -----AsnLysLeuValProValCysGluMetAlaIleLysLeuGly 306

QY 907 CAGACAAACCGC---ACCGCATTAGTACTACGCGCTCTGCGCGAAGCCACCAACATG 963
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   307 ThrAspGlyThrIleSerCysMetAlaSerTyrPheValCys----- 320

QY 964 TCGAGGACATGTTGACGGGCCATACCTACGCGCTTTGCTGGT---CGTGGTGTGATGAT 1020
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 321 -----AsnAsnIlePheAsnGlyIleMetAlaLeuAlaGlyAspThrAsnTyrTyrAsp 338
QY 1021 AWTGGGCATCCATATGATGATCCGCGACTCGGCAAGTTATTACAAC-----AAATTCCTG 1074
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   ValArgLysLysCysGluGlySerLeuCysTyrAspPheSerAsnMetGluSerPheLeu 358

QY 1075 GCAAGGACTCTGTTCATGCGACGCTATCGGCGTC---AACATCAACTACACCCAGTCCAA 1131
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   359 AsnLysLysSerValArgAspAlaLeuGlyValGlyAsnIleAspPheValSerCysSer 378

QY 1132 AATGACGCTACTACGCTTCCAGCAACAGCGGACTTTGCTGGCCCAACTTCAICGAA 1191
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   379 ProThrValTyrGlnAla-----MetLeuValAspTrpMetArg 391

QY 1192 GACCTCGAGGAGATCCTGTCTCTCCCGTGGTGTCTCC-----CTCATC 1236
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   392 AsnLeuGlu-----ValGlyIleProValLeuLeuGluAspGlyIleLysLeuVal 409

QY 1237 TAT---GGCGAGCGCGATTACATCTGCAACTGGTTCGGCGCTCAGCGGTTCCCTCGCT 1293
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   410 TyrAlaGlyGluTyrAspLeuIleCysAsnTrpLeuGlyAsnSerArgTrpValHisAla 429

QY 1294 GCGAACTACTCCCAAGCGCGCCACTTCGGAAGCGCAGGTACACGCCCTGAAGTCAAC 1353
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   430 MetGluTrpSerGlyGlnLysGluPheValAlaSerProGluValProPheIleValAsp 449

QY 1354 GCGCTCAGTATGGGAACTCGCGAGTATGTAATTTCTCTCTCATCGCTGTATGAG 1413
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   450 GlyAlaGluAlaGlyValLeuLysThrHisGlyAlaLeuGlyPheLeuLysValHisAsp 469

QY 1414 CGAGGCATGAGTCCCATCTACACCCCATCCCTCCCTCGCAATGTTTAAACCGGACT 1473
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   470 AlaGlyHisMetValProMetAspGlnProLysAlaAlaLeuGluMetLeuLysArgTrp 489

QY 1474 ATCTTCGTTGG-----GATATCGCAGAGGCGCAGAGAAGATC 1512
DB   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   490 ThrLysGlyThrLeuSerAspAlaSerAspSerGluLysLeuVal 504

RESULT 7
Q8VW01 PRELIMINARY; PRT; 429 AA.
AC Q8VW01;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Carboxypeptidase C cbp31.
GN P0434A03.7.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T.; Matsumoto T.; Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
  clone:P0434A03.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004299; BAC45113.1;
SQ SEQUENCE 429 AA; 47790 MW; 1F2F64F236475BB4 CRC64;

Alignment Scores:
Pred. No.: 1.55e-25      Length: 429
Score: 473.00      Matches: 130
Percent Similarity: 45.49%      Conservative: 77
Best Local Similarity: 28.57%      Mismatches: 178
Query Match: 16.29%      Indels: 70
DB: 10      Gaps: 16

US-09-712-338-1_COPY_55_1662 (1-1608) x Q8VW01 (1-429)
QY 154 TACTCTGGATAT-----GTCGACACTCTCCGAGTCCCATACCTTCTTCTGTTCTTC 207

```



```

Db 130 IletRpleuThrGlyGlyProGlyCysSerSerValAlaMetPheTyrGluAsnGly 149
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 CTTGCCATCAATTCGACTTTTGAUGACTACATCAACCTCACTCGTGAACAGAGTC 363
      ||||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 ProPheLysIleSerLysAspLeuSerLeuTyrTrpAsnAppheGlyTrpAspLysVal 169
      ||||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 TCCAAATTTACTATTCCTGTCCACGCAATGGSAGTCGGCTTTTCATATAGTGATAGCGTT 423
      ||||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 170 SerAsnIleLeTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAspGlu 189
      ||||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 GATGGTCCATTAACCTGTAACTACGGGGTCTCCAAAATTCGAACTTTCGAGGAGTTCAG 483
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 SerAspIleArgAsnAspGluAspGlyVal 199
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 484 GGCGGTACCAACCAATGTATGTCACCTCTGATCATCTACCAATCTTCGGCGACAGGCC 543
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 -----SerAsnAsp 202
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 544 GCTTGGAGATCCTGCAAGGATTCCTTAGTGAGTACCTAGCTAGCTAGGTCAG 603
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203 LeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro 218
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 604 TCTAAGGACATTCAGTCTATGAGCGGAGAGTATGAGGCACTATGCTCCTGCAATCTTC 663
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 LysAsnAspPhePheIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeuAla 238
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 664 AATCATTTTACGAGCAGATGAGAGAAATGCCAAACGGTAGTGTATGATGTGTTCAGCTT 723
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 SerArgValHisSerGlyAsnLysLys 253
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 724 AATTCACATCTCTGGGAATTAATAAGGCATCATCGACGAGCGATCCAGGCCCTTAC 783
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 254 AsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlnTyrGlyAla 273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 784 TACCCTGCAATTCCTGTGAACATACCTACGGTATCAAGCIGTCAACGAGACCGTCTC 843
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 TyrGlyAspTyrAlaLeuGln 289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 844 AACTACATCAAGTTTGCCAAACCAAAATGCCAAATGGTTCGACGATTTGATTTCCACTGC 903
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 GluSerLeuLys 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 904 AACAGACAAACCGCACCATTAAGTCTAGTACCTCTGCGCCGAGCCACCAACATG 963
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 AsnIleThrLysLysCysSerLeuGlyGlyGlyLeuValCysAspSerAlaValGluVal 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 964 TGCAGGACAAATGTTACAGGGCCATACCTAGCGCTTTGCTGGTGTGTATGATATT 1023
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 CysThrSerIlePheAsnLysIleValAlaLysLysSerGlyLeuAsnTyrTyrAspIle 340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1024 CGGCATCCA 1065
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 341 ArgLysLysCysValGlySerLeuCysTyrAspPhe 356
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1066 AATATTCGCAAGACTCTCTCATGAGCGTATCGGGTC---AACATCAACTACACC 1122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 IlePheLeuAsnLysGluAsnValArgLysAlaLeuGlyValGlyAspIleLysPheVal 376
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1123 CAGTCCATTAACGCTACTACCTTCCAGCAACAGCGACCTTCTCGTGGCCCAAC 1182
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 SerCysSerSerThrValTyr 389
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1183 TTTCATCGAAGACTCGAGGATCTTGTCTCTCCCGTTCGCTGTCTCC----- 1230
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 TrpMetGlnAsnLeuGlu-----ValLysIleProSerLeuValAsnAspGlyIleAsn 407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1231 ---CTCATCTAT---GGCGACGCCGATTAATCTGCAACTGGTTCGGGGTCAAGCGGT 1284
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 LeuLeuValTyrAlaGlyGluTyrAspIleLysCysAsnTrpLeuGlyAsnSerArgTrp 427
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1285 TCCCTCGCTGCAACTACTCCCAAGCCGCCAGTTCGGAAGCGAGGTTACAGCCCTG 1344
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 428 ValAspGlnMetAsnTrpSerGlyGlnLysGlyPheGlySerAlaLysAsnValSerPhe 447
QY 1345 AAGTCAACGGCGTGGAGTATGGGAAACTCGCGAGTATGTAATTTCTCCCTCACTGCG 1404
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 LeuValAspGlyLysGluAlaGlyLeuLeuLysAsnHisGlyProLeuThrPheLeuLys 467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1405 GTCTATGAGGAGGAGCAAGTCCCATCTACTACAGCCCATCGCTCCCTGCAATGTTT 1464
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 ValTyrAsnAlaGlyHisMetValProMetAspGlnProLysAlaSerLeuGlnMetLeu 487
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1465 AACCGGACTATCTTCGGT 1482
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 488 GlnAsnTrpMetGlnGly 493
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
Q9XH61
ID Q9XH61 PRELIMINARY; PRT; 501 AA.
AC Q9XH61;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Serine carboxypeptidase.
OS Matricaria chamomilla.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Anthemideae; Matricaria.
OX NCBI_TaxID=98504;
RN NCBI_TaxID=98504;
RP SEQUENCE FROM N.A.
RA Kohchi C., Yasuda H., Hirata T.;
RT "Isolation of a cDNA encoding for a carboxypeptidase, having leucine
zipper structure at the N-terminal region, from the cultured shoot
primordia of Matricaria chamomilla."
RL Plant Biotechnol. 16:409-412(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Kohchi C., Yasuda H., Hirata T.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF141384; AAD42963.2;
DR HSSP; P00729; LYSC.
DR MEROPS; S10.009; -.
DR InterPro; IPR001563; Serine_carpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR PRODOM; PD001189; Serine_carpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
KW Carboxypeptidase.
SQ SEQUENCE 501 AA; 55973 MW; DBFD8DB41860A66 CRC64;

Alignment Scores:
Pred. No.: 1-18e-24 Length: 501
Score: 461.00 Matches: 131
Percent Similarity: 44.33% Conservative: 76
Best Local Similarity: 28.05% Mismatches: 188
Query Match: 15.87% Indels: 72
DB: 10 Gaps: 15

US-09-712-338-1_COPY_55_1662 (1-1608) x Q9XH61 (1-501)
QY 115 GCAGAGGGCTTGGCAGACTACCCGGGTGCAATCTACTCTGATAT-----GTC 168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 SerAspGlyVal-----ThrValGluAspLeuGlyHisHisAlaGlyTyrTyrGlnIle 97
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 169 GACACCTCTCCGAGTCCCAIACCTTCTTCTGTTCTTCCAGCCACAGATACCCAGAA 228
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 GlnHisSerHisAlaLysMetPheTyrPhePheGluSerArgAsnAsnLysLys 117
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 229 ACTGCACCATCACATGTTGTTGAATGTTGGCCCTGGAGCGATCTTTTGATCGGTCTC 288
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 Asp---ProValIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 136
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
QY 289 TTCGAAGAGTGGGCCCTGGCATGTCATTCGACTTTTUGACTACATCAACCTCAC 348
Db 137 PheYrGluAsnGlyProPheLysIleAlaAspAsnMetThrLeuValTrpAsnGluYr 156
QY 349 TCGTGAAGAGGTCTCAAAATTTACTATCTCTGTCGCCAGCCATGGAGTCGGCTTTTCA 408
Db 157 GlyTrpAspGlnAlaSerAsnLeuIleYrValAspGlnProThrGluHisProGluPhe 176
QY 409 TATAGTACACGGTGTGATGGGTCCATTAACCTGTAACTGGGGTGTGTCGAAATTCGACG 468
Db 177 TyrSerSerAspLysArgAspIleArgHisAspGluGlnGlyVal 191
QY 469 TTTCGAGGAGTTCAGGCGCGGTACCAACCATTTGATGCCACTGTGATGATACATCAAT 528
Db 191 ----- 191
QY 529 CTTGCGCGAGAGCGCGCTGGGAGATCTCGAAGATTCTTGTAGTGACTACCTAGCTTG 588
Db 192 -----SerAspAspLeuYrAspPheLeuGlnAlaPhePheThrGluHisProGluPhe 209
QY 589 GACTCTAGGGTGCAGTCTAAGGACTTCAGTCTATGACGAGGAGAGTCTGACGGGCACTAT 648
Db 210 -----ValAspAsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyr 225
QY 649 GGTCTGTCATCTTCAATCATTTTTACGAGCAGATGAGAGAATTCGCAACGGTAGTGT 708
Db 226 IleProAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 240
QY 709 AATGGTGTTCAGCTTAATTTCAACTCTCTGGGAATTTAATACGGCATCATCGACGAGG 768
Db 241 GluGlyIleHisIleAsnLeuLysGlyPheGlyIleGlyAsnGlyLeuThrAspProAla 260
QY 769 ATCCAGGCCCTTACTACCTGATTCGCTGCTCAACATACCTACGATATCAAGGCTGTC 828
Db 261 IleGlnIleGlnAlaIleThrAspTyrAlaLeuAsp 276
QY 829 AAGCAGACCGCTCACTACATCATGAGTTTGCACCAATGCAATGTTGCCAGGAT 888
Db 277 LysGluSerGlnYr 293
QY 889 TTGATTTCCACCTGCAACAGACAAACCGC-----ACGCAATGACTGACTACGCCCTGTC 945
Db 294 AlaIleLysLeuYrGlyThrAspGlyThrValSerCysMetAlaIleValValCys 313
QY 946 CCGAAGCCACCAACATGTCGAGGACAATTTGAGGGCCATACATACGCTTGTGCTGT 1005
Db 314 -----AsnThrIlePheSerSerIleLeuSerIleAlaGly 325
QY 1006 CGT---GGTGTGATGATATTCGGCATCCA-----TATGATGACCCG 1044
Db 326 AsnIleAsnYrYrAspIleArgLysGluCysValThrSerMetCysYrAspPhe--- 344
QY 1045 ACTCCGCCAAGTATACAAATTTCTGGCAAGAGCTCTGTCATGAGCGCTATCGGC 1104
Db 345 -----SerAspMetGluThrLeuLeuAsnLysLysSerValArgGlnAlaLeuGly 361
QY 1105 GTC---AACCATCACTACCCAGTCCAAATAATGACGCTACTACGCTTTCCAGCAACA 1161
Db 362 ValGlyAspIleGluPheValSerCysSerThrThrValYrThrAlaMetLeuValVal 381
QY 1162 GCGGACTTGTCTGGCCCACTTCTGCAACACCTCGAGGAGATCCITGCTCTCCCGTG 1221
Db 382 -----LeuMetArgAsnLeuGluAlaGlyIleProGluLeuLeuGluAspGlyIle 398
QY 1222 CGTGTCTCCCTCACTATGCGGAGCGCGGATTAATCATCTGCAACTGCTGTCGGCGTCAGGCC 1281
Db 399 LysMetLeuValYrAlaGlyLysValIleCysAsnIlePheValIleValIleValIleVal 418
QY 1282 GTTTCCTCCGTCGCACTACTCCCAAGCGCGCCAGTTCCGAGCGCAGGGGTACAGGCC 1341
Db 419 TrpValHisAlaMetGluTrpAsnGlyLysGluGlnPheAsnAlaLeuSerGluAlaPro 438
```

```
QY 1342 CTGAAGTCAACGGCGTCGAGTATGGGAACTCGCGAGTATGGTAATTTCTCTCTCACT 1401
Db 439 PheGluValAspGlySerGluAlaGlyLeuLeuLysSerYrGlyProLeuSerPheLeu 458
QY 1402 CCGGCTATGAGCAGCGCCATGAAGTCCCATCTACTACCGCCATCGCTCCCTGCAATG 1461
Db 459 LysValHisAspAlaGlyHisMetValPrometAspGlnProLysAlaLeuGluMet 478
QY 1462 TTAAACCGGACTATCTTCGGT 1482
Db 479 LeuLysArgTrpMetAspGly 485
RESULT 10
Q8L6A7 PRELIMINARY; PRT; 508 AA.
AC Q8L6A7;
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Carboxypeptidase type III.
GN CP-III.
OS Theobroma cacao (Cacao) (Cocoa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Malvales; Malvaceae; Hyttnerioideae; Theobroma.
OX NCBI_TaxID=3641;
RN [1]
RP SEQUENCE FROM N.A.
RA Tazi H., McCarthy J., Bucheli P., Laloi M.;
RT "Molecular characterisation of a type-III carboxypeptidase (cacao CP-
III) from Theobroma cacao seeds.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ313514; CAC86383.1; -
DR InterPro; IPR001563; Serine_carbpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 508 AA; 56521 MW; 68B63707DC686BFE CRC64;
Alignment Scores:
Pred. No.: 1,4e-24 Length: 508
Score: 460.00 Matches: 139
Percent Similarity: 44.95% Conservative: 88
Best Local Similarity: 27.52% Mismatches: 204
Query Match: 15.84% Indels: 74
DB: Gaps: 18
US-09-712-338-1_COPY_55_1662 (1-1608) x Q8L6A7 (1-508)
QY 28 GGTAGAAGACAGCTACCAAGAACCCACCGGGGTCAAGACTCTTACAAACGCAACAAT 87
Db 65 GlyGlnValSerLeuProGluAspSerArgLeuValGluLysArgPheLysPheProAsn 84
QY 88 GTCACCATCCGCTACAAGAACCCCGGGCGACAGGGCGTCTCGGAGACTACCCCGGTGTC 147
Db 85 LeuAlaVal-----ProGlyGlyValSerVal-----GluAspLeu 96
QY 148 AATCTCTACTCTGGATAT-----GTGACACCTCTCCGAGTCCCATCTCTCTCTGG 201
Db 97 GlyHisHisAlaGlyIleYrLysLeuAlaAsnSerHisAspAlaArgMetPheYrTrp 116
QY 202 TTCCTGGAAGCCAGACATACCAACCCAGAAACTGCACCTATCATCATTTGGTGAATGGTGC 261
Db 117 PhePheGluSerArg---AsnSerLysLysAspProValIleIleIleIleIleIleIle 135
QY 262 CCGTGAAGCGGATCTTTGATCGGTCTTCTGCAAGAGTGGCGGCTGCCATGTCATTCG 321
Db 136 ProGlyCysSerSerGluLeuAlaLeuPheYrGluAsnGlyProPheThrIleAlaGlu 155
QY 322 ACTTTTGATGACTACATCAACCCCTCACCTCGTGGAGAGGCTCTCCCAATTTACTATCTCTG 381
```


Score: 439.00 Matches: 130
Percent Similarity: 42.43% Conservative: 83
Best local Similarity: 25.90% Mismatches: 191
Query Match: 15.12% Indels: 98
DB: 10 Gaps: 16

US-09-712-338-1_COPY_55_1662 (1-1608) x Q9LXH4 (1-510)

QY 16 CCGGGCGTGGTGGTAGAGACAGCTACCAAGAACCCCGGGGTCAAGACTCTTACA 75
|||||:|||||
Db 82 ProAlaAlaValAspArg----- 88
QY 76 ACCGCAACAATGTACCATCCGGTACAAAGAACCCGGGGCAGAGGGCGTCTCGAGACT 135
|||||:|||||
Db 89 -----GlySerGlyGly----- 92
QY 136 ACCCGGGGHTGCAATCTCTAC-----TCTGATAT-----GTCGACACCTCTCTCC 180
|||||:|||||
Db 93 SerProSerValGlnAspPheGlyHisAlaGlyTyrTyrLysLeuProAsnSerLys 112
QY 181 GAGTCCCATACCTCTCTCGGTCTCTCGAGCCAGACATAACCCGAAACTGCACCTATC 240
|||||:|||||
Db 113 AlaAlaArgMetPheTyrPhePheGlySerArgThrAsn---LysAlaAspProVal 131
QY 241 ACATTGTGTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTCGAAGAGTTG 300
|||||:|||||
Db 132 ValIleTyrPheThrGlyGlyProGlyCysSerSerGluLeuAlaLeuPheTyrGluAsn 151
QY 301 GGCGCTGGCATGTCATTCGACTTTGATGACTATCATCAACCTCCTCGTGAAGAG 360
|||||:|||||
Db 152 GlyProPheThrValSerAsnAsnSerSerLeuSerTyrAsnGluPheGlyTyrAspLys 171
QY 361 GTCCTCAATTTACTATTCCTGTCCTCCAGCCATGGGTGGTCTTTTCATAGTATGACG 420
|||||:|||||
Db 172 AlaSerAsnLeuIleTyrValAspGlnProValGlyThrGlyPheSerTyrThrSerAsp 191
QY 421 GTTGATGGTTCCTAATACCTGTAACCTGGGTGGTGGTGGTGGTGGTGGTGGTGGT 480
|||||:|||||
Db 192 GlnSerAspLeuArgHisAspGluAspGlyVal----- 202
QY 481 CAGGGCGGTACCCCAACCATTTGATGCCACTCTGATCGATFACCTACCAATCTGCGGAGAG 540
|||||:|||||
Db 203 -----SerAsn 204
QY 541 GCGCCCTGGAGATCCCTGCAAGGATTCCTTAGTGAGTACCTAGCTTGGACTCTAGGGTG 600
|||||:|||||
Db 205 AspLeuTyrAspPheLeuGlnAlaPhePheLysGluHisPro-----GlnPhe 220
QY 601 CAGPCTAAGGACTTCACTGATGACGAGGAGGAGTATGAGGCGCATATGTCCTGCTATC 660
|||||:|||||
Db 221 VallysAsnAspPheTyrIleThrGlyGluSerTyrAlaGlyHisTyrIleProAlaLeu 240
QY 661 TTCATCATTTTACGAGCAGATGACAGATATCCCAACGGTACTGTTAATGGTGTTCAG 720
|||||:|||||
Db 241 AlaSerArgValHisArgGlyAsnLys-----AsnLysGluGlyThrHis 255
QY 721 CTTAATTTCACTCTCTCGGAATATTAAAGGCATCATGACGAGGCGGATCCAGGCGCCT 780
|||||:|||||
Db 256 IleAsnLeuLysGlyPheAlaIleGlyAsnGlyLeuThrAsnProGluIleGlyTyrGly 275
QY 781 TACTACCTCATGCTGTGTAACAATACCTACGATATCAAGGCTGTCAACGACGCGTC 840
|||||:|||||
Db 276 AlaTyrAlaAspTyrAlaLeuAspMetAsnLeuIleThrGlnSerAspHisAspAsnLeu 295
QY 841 TACAACTACATGAGTTTGGCAACCAATGCCAATGTTGCCAGGATTTGATTTCCACC 900
|||||:|||||
Db 296 AsnArgTyr-----TyrAlaThr 301
QY 901 TCGAACAACAGAACCCGACCGCATTTAGCTGACTACGCC-----CTCTGCGCCGAA 951
|||||:|||||
Db 302 CysGlnGlnSerIleLeuGlySerAlaAspGlyGlyGluGlyAspAlaCysAlaSer 321
QY 952 GCCACCAACATGTGCGAGGACAAATGTTGAGGGGCCATACTAGCGCTTTGCTGGTGGT--- 1008

Db 322 SerTyrThrValCys---AsnAsnIlePheGlnLysIleMetAspIleAlaGlyAsnVal 340
|||||:|||||
QY 1009 GGTGTGTATGATATTCGGCATCCATATGATGACCCGACTCCGCCAAGITATTACAAAC--- 1065
|||||:|||||
Db 341 AsnTyrTyrAspValArgLysGlnCysGluGlySerLeuCysTyrAspPheSerAsnMet 360
QY 1066 ---AAATTTCTGGCAAGGACTCTGTCATGAGGAGCTATCGCGGTC---AACATCACTAC 1119
|||||:|||||
Db 361 GluAsnPheLeuAsnGlnLysSerValArgLysAlaLeuGlyValGlyAspIleGluPhe 380
QY 1120 ACCAGTCCAAATATGACGCTACTACGCTTTCACGCAAAACAGGCGACTTGTCTCGGCC 1179
|||||:|||||
Db 381 ValSerCysSerThrAlaValTyrGluAlaMetGlnMetAsp-----TyrMet 396
QY 1180 AACTTTCATCGAA---GACCTCGAGGAGATCCTTCTCTCCCGTGGTGTCTCCCTCATC 1236
|||||:|||||
Db 397 ArgAsnLeuGluValGlyIleProAlaLeuLeuGlnAspGlyIleLysLeuValTyr 416
QY 1237 TATGGGAGCGGATTCATCTGCAACTGGTTCGGCGGTCAGCGCGTTTCCCTCGCTGCG 1296
|||||:|||||
Db 417 AlaGlyGluTyrAspLeuIleCysAsnTyrLeuGlyAsnSerLysTyrValHisGluMet 436
QY 1297 AACTACTCCCAACGCGCCAGTTCGGAAGCGGAGGTACACGCGCTGAAAGTCAACGCGC 1356
|||||:|||||
Db 437 GluTyrSerGlyGlnLysGluPheValAlaAlaAlaThrValProPheHisValAspAsn 456
QY 1357 GTCGAGTATGGGAAATCTCGAGTATGTTGTAATTCCTTCTACTCGCGTCTAGGGCA 1416
|||||:|||||
Db 457 LysGluAlaGlyLeuMetLysAsnTyrGlySerLeuThrPheLeuLysValHisAspAla 476
QY 1417 GGCGATGAATCCCTACTACCGCCATCGCTCCCGCAATGTGTTAAACCGGAGATATC 1476
|||||:|||||
Db 477 GlyHisMetValProMetAspGlnProLysAlaAlaLeuGlnMetLeuGlnAsnTyrMet 496
QY 1477 TTCGGT 1482
Db 497 GlnGly 498

RESULT 12
Q93ZC3 PRELIMINARY; PRT; 510 AA.
ID Q93ZC3
AC Q93ZC3, 2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE AT3945010/F14D17.80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A.
RA Cheuk K., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam J.M., Lee J.M., Lin S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY057639; AAL15270.1;
DR InterPro; IPR001563; Serine_carbpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; Serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXPTASEC.
DR PRODOM; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 510 AA; 56837 MW; 704BC00C4E64BF08 CRC64;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

DR PRINTS; PR00724; CRBOXYPTASEC.
 DR PRODOM; PD001189; Serine_carbpept; 6.
 DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 4.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 2.
 SQ SEQUENCE 2105 AA; 234335 MW; 139CE6E96D75B3FB CRC64;

Alignment Scores:
 Pred. No.: 4,92e-21 Length: 2105
 Score: 413.50 Matches: 150
 Percent Similarity: 38.52% Conservative: 85
 Best Local Similarity: 24.59% Mismatches: 202
 Query Match: 14.24% Indels: 173
 DB: 5 Gaps: 26

US-09-712-338-1_COPY_55_1662 (1-1608) x Q17679 (1-2105)

```

QY 10 AGTACACGGGTCGTCGGTAGAAGACAGTACCACCAAGAACCCACGGGGTCAAGACT 69
D 1521 SerThrProAlaGlyIle-----AspValThrProGlnProLeuThrAsn 1535
QY 70 CTTACAACCGCAACAATGTCAACCATCCGGTACAAAGGAACCCGGGCGAGGGCGTCTGC 129
D 1536 ValIleSerGlnThrAsnCysThr-----ThrGlyGlnThrAspArgIle 1550
QY 130 GAGACTACCCCGGGTGTC-----AAATCCTACTCTGGATATGTGAC 171
D 1551 IleAsnLeuProGlyLeuProAlaAspMetLeuPheLysGlnTyrSerGlyPheLeuAsp 1570
QY 172 ACCTCTCCGAGTCCCAATACCTTCTTCTGGTCTTCGAAGCCACACATAACCCAGAACT 231
D 1571 GlyLeuSerGlyHisLysValHisTyrTrpLeuValGluSerGluAsnAsnProSerThr 1590
QY 232 GCACCTATCATGTTGGTGAATGGTGGCCCTGGAAGCGATTTCTTGATCGGTCTCTTC 291
D 1591 AspProLeuLeuTrpLeuAsnGlyGlyProGlySerSerSerLeuMetGlyLeuPhe 1610
QY 292 GAAGAGTGGGCCCTGCCATGTCAATTCGACTTTTGATGACTACATC---AACCCCTCAC 348
D 1611 GluGluAsnGlyProPheArgValSerLysAspSerGlnThrLeuSerArgAsnProTyr 1630
QY 349 TCGTGGACGAGGCTCCAAATTACTATCTCTGTCGCCAGCCATTCGGAGTCGGCTTTTCA 408
D 1631 SerTrpAsnLysPheAlaAsnValLeuTyrLeuGluSerProIleGlyValGlyTyrSer 1650
QY 409 TATAGTGATACGGTTGATGGGTCCATTAACCTCTAATCGGGTCTGCGAAATTCGAGC 468
D 1651 Tyrala----- 1652
QY 469 TTTCGAGAGTTACGGCCCGGTACCCAAACCATGATGCCACTCTGATCGATPACTACCAAT 528
D 1653 -----TyrAsnAsnThrAsnIleGlnTyrAspValThr--- 1664
QY 529 CTTCGCCGAGCGCGCTGGAGATCTCTGAAGANTTCCTTAGTGGACTACCTAGCTTG 588
D 1665 ---ThrAlaGlnGluAsnTyrAlaAlaLeuLysSerPhePheAlaGlnTyrPro----- 1681
QY 589 GACTCTAGGGTGCAGCTTAAGCACTTCACTATGTCGAGGAGAGATGATGAGGCGACVAT 648
D 1682 -----GlnTyrThrThrSerAspPheTyrThrThrGlyGluSerTyrAlaGlyValTyr 1699
QY 649 GGTCTCGCATCTTCATCATTTTTACAGCAGAGATGAGAAATGCCAAGGTAGTGT 708
D 1700 LeuProGlyLeu-----SerAlaLeuValGlnGlyIleLys 1712
QY 709 AATGGT---GPTCAGCTTAATTTCACTCTCTGGGANTTATTAAAGGCATCATCGACGAG 765
D 1713 SerGlyAspIleAsnIleAsnTyrLysGlyValSerIleGlyAsnGlyValIleAspLys 1732
QY 766 -----GCGATCCAGCGCCCTTACTACCTCGAATTCGCTGTAACAAT 807
D 1733 ArgThrAspMetAsnSerGlnLeuHisTyrGlnTyr----- 1745
QY 808 ACCTACGGTATACAGGCTGTCAACGAGACCGCTCTACAAC----- 846

```

```

Db 1746 ---HisGly-----GlyIleSerAlaThrThrTyrAsnThrAlaLeuGlnLeuCysCys 1762
QY 847 -----TACATGAAGCTTTTCCCAACCAATG 870
D 1763 SerGlyAspGluPheLysCysArgPheSerAspArgMetThrAsnPheAsnSerIle 1782
QY 871 CCAATGGT-----TGCAGGAT---TTGATTTCCACTCCAAACAGACA 912
D 1783 ProTrpGlyAspLeuSerAspProCysTyrAspPheValValAlaThrGlyAlaGlnLeu 1802
QY 913 AACCGACCGCATTAGCTAGCTAGCCCTCTGCGCGGAGAACCAACATGTGACGGAC 972
D 1803 LeuLeuThrAlaPheAspProTyrAsnMetTyrGlnGlnCysTrpThrIle----- 1819
QY 973 AATGTTGAGGGCCATACACTACCCCTTTCGTGCTGCTGTGTATGATATTCGGCATCA 1032
D 1820 -----Pro 1820
QY 1033 TATGATGACCGACTCCGCCAAGTTATTACACAAA----- 1068
D 1821 TyrAsnAspThrThrProArgThrProTyrGlyGlnThrTrpThrGlyLeuAsnTyrGlu 1840
QY 1069 -----TITCTG 1074
D 1841 SerSerAspPheLeuAsnGlyTyrProCysTyrAspSerAlaMetGluGlyTyrLeu 1860
QY 1075 GCAAGGACTCTGTCTATGAGCTATCGCGTC-----AACATCAACTACACCCAGTCC 1128
D 1861 AsnArgProValValArgLysAlaLeuAsnIleProAspSerValProTyrTrpAlaAla 1880
QY 1129 AATAATGACGCTACTAGCTTTCAGCAAAACAGGCGACTTTGTCTGCCCAACTTCATC 1188
D 1881 AsnAsnAsnIleIleAsnAlaTyrAsnGlnGlnValAspSerIleValProAsn----- 1898
QY 1189 GAAGACTCGGAGAGATCTTCTGCT-----CTCCCGCTGCTGCTCTCCCTCATCTAT 1239
D 1899 -----LeuGlnIleIleMetAlaAsnAlaProAlaAsnPheLysMetLeuTyrSer 1916
QY 1240 GCGCAGCGCATATACATCTGCAACHTGTCGCGGTCCAG----- 1278
D 1917 GlyAlaAlaAspThrMetValAsnTrpLeuGlyAlaGluIlePheThrAlaAsnAsnPhe 1936
QY 1279 ---GCGCTTTCCTCGCTGCGAATCTACCCAAAGCGCC-----CAGTTCGGAAGC 1326
D 1937 AlaAlaLeuGlyLeuThrThrSerSerProArgAlaGlnTrpThrPheGlnTyrAsnSer 1956
QY 1327 GCAGGTACACGCCCTGAAAGTCAACGGCTCGAGTATGGGGAACCTCGCGAGTATG 1386
D 1957 Thr---PheGlnPro---ThrValAlaGlyTyrGlnThrSerTyrThrSerAsnAlaIle 1974
QY 1387 AATTTCTCTCCTCAGCTGCTATGAGCGAGCCCATGAAGTCCCATCTACCGACCCATC 1446
D 1975 AsnIleAspValLeuThrValLysGlySerGlyHisPheValProLeuAspArgProGln 1994
QY 1447 GCTCCCTCGCAATGTTTAACCGGACTATCTTCGTTGGTATGATCGAGAGGCGCAGAG 1506
D 1995 GlnAlaLeuGlnMet-----IleTyrAsnPheValLysSerArgGlyTyrAsn 2010
QY 1507 AAGATCTGGCCAGCTACAGACGAATGAGCGGTACAGCTACGATACGATACAGCTGCTCC 1566
D 2011 ThrProPheAspLeuAsnSerAsnPheThrThrThrThrThrThrThrThrThrThr 2030
QY 1567 GTCCG-----CTGCCCTACGGCTACCGAC 1590
D 2031 ThrProGlyThrGlyProThrValThrAla 2040

```

RESULT 14

Q9FMX9 PRELIMINARY; PRT; 499 AA.
 AC Q9FMX9;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Serine carboxypeptidase II-like protein.
 DE Arabidopsis thaliana (Mouse-ear cross).
 OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 SEQUENCE FROM N.A.
 RP RP
 RC STRAIN=Columbia;
 RX MEDLINE=98162728; PubMed=9501957;
 RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
 RT Sequence features of the regions of 1,191,918 bp covered by seventeen
 RT physically assigned pl clones.";
 RL DNA Res. 4:401-414(1997).
 DR EMBL: AB007648; BAB11176.1; -
 DR HSSP: P08819; LWHT.
 DR MEROPS: S10.005; -.
 DR InterPro: IPR001563; Serine_carbpept.
 DR InterPro: IPR000379; Serestr.site.
 DR Pfam: PF00450; serine_carbpept; 1.
 DR PRINTS: PR00724; CRBOXYPTASEC.
 DR ProDom: PD001189; Serine_carbpept; 2.
 DR PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
 DR PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
 DR Carboxypeptidase.
 SK SEQUENCE 499 AA; 56329 MW; 6150FB81A2E7BCBD CRC64;
 QW

Alignment Scores:
 Pred. No.: 9,67e-21 Length: 499
 Score: 406.50 Matches: 137
 Percent Similarity: 41.17% Conservative: 75
 Best Local Similarity: 26.60% Mismatches: 196
 Query Match: 14.00% Indels: 107
 DB: 10 Gaps: 22

US-09-712-338-1_COPY_55_1662 (1-1608) x 09FMX9 (1-499)

QY	97	CGGTACAAGAA---CCCGGGCAGAGGGCGTCGCGAGACTACCCGGGTGCACA---150
Db	51	ArgValLysGluLeuProGly-----GlnProProValLysPhe63
QY	151	----TCCTACTGTGGATGTGCAGACCTCTCCCGAGTCCCAT-----ACCTTCTTC198
Db	64	ArgGlnTyrAlaGlyTyrVal---ThrValAsnGluThrHisGlyArgAlaLeuPheLyr82
QY	199	TGTTTCTTCGAAGCAGACATCAACCCAGAACTGCACATTCACATTCGTGGTGAATGGT258
Db	83	IrpPhePheGluAlaThrGluAsnProSerLysLysProValLeuLeuTrpLeuAsnGly102
QY	259	GGCCCTGGAAGCGATTCCTTTC---ATCGGTCTCTTCAAGAGTGTGGCCCTTCGCATGTC315
Db	103	GlyProGlyCysSerSerIleGlyPheGlyAlaAlaGluLeuGlyProPhePhePro122
QY	316	AATCGACTTTTGTAGAC-----TACATCAACCTCACTCGTGGACAGAGGTCTCCAA369
Db	123	GlnAsnSerSerGlnProLysLeuLysLeuAsnProTyrSerIrpAsnLysAlaAlaAsn142
QY	370	TTACTATTCCTGTCGCCAGCCATTCGGAGTCCGGCTTTTCATATAGTATAGCGGTGATGG329
Db	143	LeuLeuPheLeuGluSerProValGlyValGlyPheSerTyrThrAsnThr-----159
QY	430	TCCATTAAACCCGTGAAGTGGGTGCTGCAGAAATTCGAGCTTCGAGGAGTTCAGGCCGG489
Db	159	-----159
QY	490	TACCAACCAATTGATGCCACTCTGATCGATACCTACCAATCTGCCCGAGAGCCGCTGG549
Db	160	-----SerArgAspIleLysGlnLeuGlyAspThr---valThrAlaArgAspSerTyr176
QY	550	CAGATCTGCAGGATTCCTTAGTGGACTACCTAGCTGTGACTCTCAGGTGCGAGTCAAG609

Wed Sep 17 15:23:15 2003

us-09-712-338-1_copy_55_1662.n2p.rspt

Page 17

Db 451 LenPhe 452

Search completed: September 16, 2003, 18:46:58
Job time : 138.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:47:04 : Search time 61 seconds
(without alignments)
8368.273 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662

Perfect score: 536

Sequence: 1 cttccagggaagtacaccgcg.....gcattgtccagtggtgtatg 1608

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Word size: 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09712338/runat_16092003_144403_14541/app.query.fasta_1.1799
-DB-A.Geneseq_19Jun03 -QFMT=fastan -SUFFIX=n2poli.rag -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR_MIN=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09712338 @CGN_1_1_92.runat_16092003_144403_14541 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A.Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	535	99.8	554	19 AAW56099	A. oryzae ATCC2038

ALIGNMENTS

RESULT 1

AAW56099
ID AAW56099 standard; Protein; 554 AA.

XX AAW56099;

XX 27-AUG-1998 (first entry)

XX A. oryzae ATCC20386 carboxypeptidase I protein.

XX Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous; food industry.

XX Aspergillus oryzae.

XX Key Location/Qualifiers

FT Peptide 1..18

FT /label= signal

FT Protein 19..554

FT /label= carboxypeptidase I

XX WC9814599-A1.

XX 09-APR-1998.

XX 03-OCT-1997; 97WO-US17977.

XX 27-NOV-1996; 96US-0757534.

XX 04-OCT-1996; 96US-0726880.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO-NORDISK AS.

XX Barka R, Blinkovsky A, Brown K, Dambmann C, Golightly E;

XX Klotz A, Mathisen TE, Rey M;

XX WPI; 1998-240098/21.

XX N-PSDB; AAV28620.

XX Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous substrates, useful for improving flavour of foods

XX Claim 1; Fig 3; 82pp; English.

XX This sequence represents carboxypeptidase I from Aspergillus oryzae. This polypeptide has an optimal activity in the range of pH 3.0-7.5 at 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a residual activity of at least 65.5% after 30 minutes at pH 4.0 and 60 deg. C. It also has the capacity to hydrolyse X from N-CBZ-Ala-X where N-CBZ is N-carboxy and X is any amino acid. The carboxypeptidases can be used for obtaining hydrolysates (which can be enriched in free glutamic acid or peptide bound glutamic acid residues) from proteinaceous substrates. The carboxypeptidases can be used in flavour-improving compositions in the food industry. The products can also be used for the production of polypeptides free of carboxypeptidase activity.

XX Sequence 554 AA;

XX Alignment Scores:

Pred. No.: 0 Length: 554

Score: 535.00 Matches: 536

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00%
 Query Match: 99.81%
 DB: 19 Indels: 0
 Gaps: 0

US-09-712-338-1_COPY_55_1662 (1-1608) x AAW56099 (1-554)

QY	1	CTTCCAGGAGTACACCGCGTCCGTCGGTAGAAGACAGTACCCAGAACCCACCGGG	60
Db	19	LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly	38
QY	61	GTCACAGCTTACACCGCGAACAATGTCACCATCCGGTACAGAACCCCGGGCAGAG	120
Db	39	ValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu	58
QY	121	GGGCTCTGGAGACTACCCCGGTGTCAATCTACTCTGGATATGTCGACACTCTCC	180
Db	59	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyThrValAspThrSerPro	78
QY	181	GAGTCCCATACCTTCTCTGGTTCCTTCGAAGCCAGACATAACCCAGAACTGCACCTATC	240
Db	79	GluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIle	98
QY	241	ACATTTCTGGTGAATGGTCCCTCGAAGCGATTCTTGTATCGTCTCTTCGAAGAGTTG	300
Db	99	ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu	118
QY	301	GGCCCTTGCCATGCAATTCGACTTTTGATGACTACATCAACCTCCTCACTCGTGGACGAG	360
Db	119	GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu	138
QY	361	GTCCTCAATTTACTATTCCTGTCCAGCCATTCGGAGTCGGCTTTTCATATAGTAGTACG	420
Db	139	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr	158
QY	421	GTTCATGGGTCCATTAACCTGTAACTGGGGTCCGTCGAAATTCGAGCTTTCGAGAGTT	480
Db	159	ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal	178
QY	481	CAGGGCGGTPACCAACCATGTATGCCACTCTGATCATACTACCAATCTTGGCCGACAG	540
Db	179	GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu	198
QY	541	GGCGCTTGGAGATCTCTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGTG	600
Db	199	AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal	218
QY	601	CAGTCTAAGGACTTCAGTCTATGGACGAGAGGTATGGAGGCACATGCTCGCATTC	660
Db	219	GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe	238
QY	661	TTCAATCATTTTACGACGAGAATGAGAGAATGCCAACGGTAGTGTATATGGTGTTCAG	720
Db	239	PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln	258
QY	721	CTTAATTTCACTCTCTGGGAATTTAAGGGCAATCATCGACAGGCGCATCGCCGCCCT	780
Db	259	LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro	278
QY	781	TACTACCTCAATTCGCTGTGCAACAATACCTACGGATCAAGGCTGTCAACGACGACGTC	840
Db	279	TyrTrpProGluPheAlaValAsnAsnThrIleGlyIleLysAlaValAsnGluThrVal	298
QY	841	TACAACCTACATGAGTTTGGCCACCAATGCCAATGGTTGCCAGGATTTGATTTCCACC	900
Db	299	TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr	318
QY	901	TGCAACAGACAAACCCCGCATTTAGCTGACTACGCCCTCTCGCCGCGAAGCCACCAAC	960
Db	319	CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn	338
QY	961	ATGTGACGGGACAATTTGAGGGGCCATACCTACGCCCTTGTGCTGGTCTGGTGTGTATCAT	1020
Db	339	MetCysArgAspAsnValGluGlyProTyrThrAlaPheAlaGlyArgGlyValTyrAsp	358

Search completed: September 16, 2003, 18:58:12

Job time : 63 secs

QY	1021	ATTTCGCATCCATATGATGACCCGACCTCCGCCAGTATTATACAACTTCCTGGCAAG	1080
Db	359	IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys	378
QY	1081	GACTCTGTCTATGACCGCTATCGGCTCAACATCACTACACCCAGTCCCAATAATGAGTTC	1140
Db	379	AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspVal	398
QY	1141	TACTAGCTTTCCAGCAACACAGCGGACTTTGTCTGGCCCACTTCATCGAAGACTCGAG	1200
Db	399	TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu	418
QY	1201	GAGATCCTTCTCTCCCGCTCGCGTCTCCCTCATCTATGCGGACGCGCGATACATCTGC	1260
Db	419	GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys	438
QY	1261	AACCTGGTTCGGCGTACGCGGCTTCCCTCGCTCGGAACCTACTCCCAAGCGCCCACTTC	1320
Db	439	AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe	458
QY	1321	CGAAGCGCAGGGTACACGCGCCCTGAAAGTCAACGCGCTCGAGTATGGGAAACCTCGCAG	1380
Db	459	ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu	478
QY	1381	TATGGTAATTTCTCTTCACTCGCGTCTATGAGCAGGCCCATGAAGTCCCATACCTACCAG	1440
Db	479	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	498
QY	1441	CCCATCGCCCTCCCTGCAATTGTTTAAACCGGACTATCTTCGTTGGGATATCGCAGAGGC	1500
Db	499	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly	518
QY	1501	CAGAAGAAGATCTGGCCAGCTACAGACGAATGGAACGGCTACAGCTACGCATACACAG	1560
Db	519	GlnLysLysIleIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln	538
QY	1561	TGTPCCGTGGCGTGGCTGACGGTACCGCATGCCAGCATGCCAGTGTGGTATG	1608
Db	539	SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet	554


```

Db      119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu 138
      361 GICTCCAAATTACTATTCCTGTCCTCCAGCCATTGGAGTCGGCTTTTCATATAGTATACG 420
Db      139 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 158
QY      421 GTTGATGGTCCATTAAACCTCTTAACCTGGGGTGGCTGCAGAAATTCGAGCTTTCAGAGATT 480
Db      159 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 178
QY      481 CAGGGCCGTACCAACCATTTGATGCTACTCTGATGATACATACCAATCTTCCTCCGAGAG 540
Db      179 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGlu 198
QY      541 GCCGCTGGGAGATCCTCAAGATTCCTTACTGGACTACCTAGCTTGGACTCTAGGGTG 600
Db      199 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 218
QY      601 CAGTCTAAGGACTTCAGTCTATGAGGAGAGACTATGAGGAGCACTATGGTCTCGCATTC 660
Db      219 GlnSerLysAspPheSerLeuThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238
QY      661 TTCATATCTTTTACGACGAGAATGAGAGAATTCGCAAGGTAGTGTAAATGGTGTTCAG 720
Db      239 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 258
QY      721 CTTAATTTCACTCTCTGGGAATTAATTAACGGCATCATCGACGAGGGGATCCAGGCCCT 780
Db      259 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaPro 278
QY      781 TACTACCTGATTCGCTGTGCAACAATACCTACGGTATCAAGGTGTCAAGGAGACCGGTC 840
Db      279 TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal 298
QY      841 TACAACATACATGAAGTTTGGCAACCAATGTCGAATGTTGCCAGGATTTGATTTCCACC 900
Db      299 TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 318
QY      901 TGCACACAGACAACCCGACCGCATTAGCTGACTACGCCCTCTCGCGGAGAGCCACCAAC 960
Db      319 CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 338
QY      961 ATGTGACGAGGACAAATGTAGGGGCCATACTACGCCCTTCTGCTGCTGGTGGTGTATCAT 1020
Db      339 MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 358
QY      1021 ATTCGGCATCCATATGATACCGGACTCCGCGCAAGTTATTACACAAATTTCTGGCAAG 1080
Db      359 IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 378
QY      1081 GACTCTGTATGACCGGTATCGGGGTCAACATCACTACACCCAGTCCCAATATGAGGTC 1140
Db      379 AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 398
QY      1141 TACTAGCGCTTTCAGCAAAACAGCGCTTTGCTGGCCCACTTCATCGAAGACCTCGAG 1200
Db      399 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 418
QY      1201 GAGATCCTTGCTCTCCCGTGGCTGTCTCCCTCATCTATGGCGAGCGGATTTACATCTGC 1260
Db      419 GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspIleCys 438
QY      1261 AACTGGTTCGGGGTTCAGCCGTTTCCCTGCTGGCACTACTTCCCAAGCCGCCAGTTC 1320
Db      439 AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 458
QY      1321 CGAAGCGCAGGTACACGCCCTGAAAGTCAACGGGCTGAGTATGGGGAACCTCCGAG 1380
Db      459 ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyThrArgGlu 478
QY      1381 TATGTATTTCTCCTCACTCGCTCTATGAGGAGGCGCATCAAGTCCCATCTACGAG 1440

```

Search completed: September 16, 2003, 19:04:29
Job time : 24 secs

```

Db      479 TyrGlyAsnPheSerPheThrArgValTyrCluAlaGlyHisGluValProTyrTyrGln 498
QY      1441 CCCATCCCTCCCTGGCAATTGTTAACCGGACTATCTTCGGTTGGGATATCGCAGAGGC 1500
Db      499 ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 518
QY      1501 CAGAAGAAGATCTGGCCACGCTACAGACCAATGGACGGCTACAGCTACGCATACACAG 1560
Db      519 GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrHisThrGln 538
QY      1561 TCCTCGTGGCGCTGCTACGGCTACGAGCATGTCCAGTGTGGTATG 1608
Db      539 SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet 554

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

No matches found

Search completed: September 16, 2003, 19:06:27
Job time : 55 secs

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:59:05 : Search time 55 Seconds
(without alignments)
8706.201 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 536
Sequence: 1 cttccagaagtacaccgc.....gcattccagttgttatg 1608

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 556269 seqs, 148893369 residues

Word size: 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09712338/runat_16092003_144405_14644/app_query.fasta_1.1799
-DB=Published Applications AA -QFMT=fastan -SUFFIX=n2poli.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=100
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09712338.cgn_1_1_15@runat_16092003_144405_14644
-NCPG=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications_AA:

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:56:25 : Search time 36.5 Seconds

(without alignments)
8473.383 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662

Perfect score: 536

Sequence: 1 ctccagggaagtacacggc.....gcattgccagtgttggtatg 1608

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Word size: 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPRO.spool_P/US09712338/runat_16092003_144405_14586/app_query.fasta_1.1799
-DB=PIR_76 -QMT=fastan -SUFFIX=n2poli.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cgi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338_@CGN_1_1_57_@runat_16092003_144405_14586 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%
No.	Score	Match Length DB ID Description

No matches found

Search completed: September 16, 2003, 19:03:33
Job time : 36.5 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:48:44 : Search time 20.5 Seconds
(without alignments)
7377.458 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 536
Sequence: 1 cttccaggaggtacaccgc.....gcattccaggtgtgtgtatg 1608

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Word size: 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO_scol_p/US09712338/runat_16092003_144404_14552/app_query.fasta_1.1799
-DB=SwissProt_41 -FWT=fastan -SUFFIX=n2poli.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAIRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338.cgn_1.1.24.@runat_16092003_144404_14552 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description

No matches found

Search completed: September 16, 2003, 18:59:01
Job time : 20.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 16, 2003, 18:49:49 ; Search time 90 Seconds
(without alignments)
9221.080 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662

Perfect score: 536

Sequence: 1 cttccaggaagtacaccggc.....gcattgctcagtgattggtatg 1608

Scoring table:

OLIGO	Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 830525 seqs, 258052604 residues

Word size: 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_P/US09712338/runat_16092003_144404_14563/app_query.fasta_1.1799
-DB=SPTREMBL_23 -OFMT=fastan -SUFFIX=n2poli.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 -CGN_1_1_123 -runat_16092003_144404_14563 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1:	sp_archaea:
2:	sp_bacteria:
3:	sp_fungi:
4:	sp_humani:
5:	sp_invertebrate:
6:	sp_mammal:
7:	sp_mhc:
8:	sp_organelle:
9:	sp_phase:
10:	sp_plant:
11:	sp_rodent:
12:	sp_virus:
13:	sp_vertebrate:
14:	sp_unclassified:
15:	sp_rvrius:
16:	sp_bacteriap:
17:	sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	434	81.0	555	3 Q96VZ9	Q96VZ9 aspergillus

ALIGNMENTS

RESULT 1

Q96VZ9
ID Q96VZ9 PRELIMINARY: PRT: 555 AA.
AC Q96VZ9;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Carboxypeptidase S1 precursor (EC 3.4.16.6).
GN CPT.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
RX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FK3;
RA van den Broek P.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF394242; AAK77166.1; -
DR InterPro: IPR001563; Serine_carbpept.
DR Pfam: PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPEPTASEC.
DR ProDom: PD001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
KW Carboxypeptidase; Hydrolase; Signal.
FT SIGNAL 1 18
FT SIGNAL 18
SQ SEQUENCE 555 AA: 61168 MW: 456563H0CEB55222 CRC64;

Alignment Scores:

Pred. No.: 0 Length: 555
Score: 434.00 Matches: 535
Percent Similarity: 99.81% Conservative: 0
Best Local Similarity: 99.81% Mismatches: 1
Query Match: 80.97% Indels: 0
D: 3 Gaps: 0
US-09-712-338-1_COPY_55_1662 (1-1608) x Q96VZ9 (1-555)

QY	1	CTTCAGGAGTACACCGCGTCCGTAGAGACAGCTACCAAGAACCCACCGG	60
Db	19	LeuProGlySerIhrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGly	38
QY	61	GTCAGAGACTTACAAACGCAACAAATGTCACCATCCCGGTACAGGAACCCGGG	120
Db	39	VallysThrLeuThrIhrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu	58
QY	121	GGCGTCTCGAGACTACCCCGGGTGTCAAATCTTCTGATATGTGACACCTCTCC	180
Db	59	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro	78
QY	181	GAGTCCCATACCTCTCTGTTCTTTCGAAGCCAGACATAACCCAGAACTSCACCTATC	240
Db	79	GluSerHisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIle	98
QY	241	ACATTGTGGTTGAATGTGGCCCTGGAAGCCATTCTTGTATCGGTCGTCGAGAGTTG	300
Db	99	ThrLeuIrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu	118
QY	301	GGCGCTTCCCATGTCAATTCGACTTTTGATCATTACATCAACCCCTACTCGTGAACGAG	360
Db	119	GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu	138
QY	361	GTCTCCCAATTACTATTCTCTGCCAGCCATTCGGAGCTCGGCTTTTCATATGATACG	420
Db	139	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerIyrAsnAspThr	158
QY	421	GTTCATGGGTCCATTAACCTGTAACTGGGGTGTGCGTGGAAATTCGACGCTTTCGAGGAGTT	480
Db	159	ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal	178

Search completed: September 16, 2003, 19:02:12
Job time : 92 secs

QY	481	CAGGCGCGGTACCCAAACCATTTGATGCCACTCTGATGCGATACACAAATCTTGGCCGACAG	540
DB	179	GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGln	198
QY	541	CGCGCTTGGAGATCCTGCAAGGATTCCTTACTGGACTACCTAGCTTGGACTCTAGGGTG	600
DB	199	AlaAlaTrpGlnLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal	218
QY	601	CAGCTCAAGGACTTCAGTCTATGGCGGAGAGCTATGGAGGSCACTATGGTCCGCAATC	660
DB	219	GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe	238
QY	661	TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCACGGTAGTGTTAATGGGTTCAG	720
DB	239	PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln	258
QY	721	CTTAATTTCAACNCTCTGGGAATTTAATACGGCATCATCGACGAGCGCATCCAGGCCCT	780
DB	259	LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro	278
QY	781	TACTACCTCGAATTCGCTGTGGAACAATACCTACGGTATCAAGCGTGTCAACGAGACCGTC	840
DB	279	TyrTyrProGlnPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGlnThrVal	298
QY	841	TACAACATACATGAAGTTTGCCACCAACAAATGCCAAATGGTTGCCAGGATTTGATTTCCACC	900
DB	299	TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr	318
QY	901	TGCAACACAGACAACCGCACCGCATTAGCTAGCTACGCCCTCTGGCCGAGCCACCAC	960
DB	319	CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn	338
QY	961	ATGTGCGAGGACAATGTTGAGGGCCCNACTAGCCCTTTGCTGGTGGTGGTGTGTATGAT	1020
DB	339	MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp	358
QY	1021	ATTCCGGCATCCATATGATGACCGCACCTCCCAAGTTATTACAACAAAATTTCTGGCAAG	1080
DB	359	IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys	378
QY	1081	GACTCTGTCATGGACCGTATCGGCGTCAACATCACTACACCCAGTCCCAATATGACGTTC	1140
DB	379	AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspVal	398
QY	1141	TACTACGCTTTCAGCAAAACAGGCGACTTGTCTGGCCCAACTTCATCGAAGACCTCGAG	1200
DB	399	TyrTyrAlaPheGlnGlnThrThrGlyAspPheValTrpProAsnPheIleGlnAspLeuGln	418
QY	1201	GAGATCTCTGCTCTCCCGTGGTGTCTCCCTCATCTATGGCGACCGCATACATCTCG	1260
DB	419	GlnIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys	438
QY	1261	AACTGGTTTCGGCGGTACAGCGGTTTCCTCGCTCGGAACCTACCTCCCAAGCGGCCAGTTC	1320
DB	439	AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPhe	458
QY	1321	CGAAGCGCAGGTACACGCCCTCGAAAGTCAACGGCGTCGAGTATGGGAAATCTCGCAG	1380
DB	459	ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu	478
QY	1381	TATGTAATTTCTGTTTCACCTCGCGGTCTATGAGGCGAGGCCCATGAHCCCATCTACCG	1440
DB	479	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln	498
QY	1441	CCCATCGCTCCCTCGCAATTCGTTAAACGGGACTATCTCGGTTCGGGATATCGCAGAGGC	1500
DB	499	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly	518
QY	1501	CAGAAGAAGATCTGGCCCGAGCTACAAGACGAATGGAACGGCTACAGCTACGATACACAG	1560
DB	519	GlnIstLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrAsnThrGln	538

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:03:35 : Search time 7741 Seconds
(without alignments)
8497.956 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 1608
Sequence: i cttccaggaagtacaccgc.....gcattgtccagtgtgtgtatg 1608

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 100

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1608	100.0	1662	6	ARI29928	ARI29928 Sequence
2	531	33.0	2245	8	AF394242	AF394242 Aspergill

ALIGNMENTS

RESULT 1
ARI29928
LOCUS ARI29928 1662 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6187578.
ACCESSION ARI29928
VERSION ARI29928.1 GI:14117825
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1662)
AUTHORS Blinkovsky,A., Berka,R., Rey,M., Golightly,E., Klotz,A.,
Mathisen,T.Erik., Dambmann,C. and Brown,K.M.
TITLE Carboxypeptidases and nucleic acids encoding the same
JOURNAL Patent: US 6187578-A 1 13-FEB-2001;
FEATURES Location/Qualifiers
source
1..1662
BASE COUNT 396 a 468 c 398 g 400 t
ORIGIN
Query Match 100.0%; Score 1608; DB 6; Length 1662;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CTTCCAGSAGTACACCGGCTCGTGGTAGAGACAGCTACCCAGAACCCACCGGG	60
Db	55	CTTCCAGSAGTACACCGGCTCGTGGTAGAGACAGCTACCCAGAACCCACCGGG	114
Qy	61	GTCAGACTCTTACAAACCGCAACAATCTCAACCTCGGTACAGAACCCGGGCGAG	120
Db	115	GTCAGACTCTTACAAACCGCAACAATCTCAACCTCGGTACAGAACCCGGGCGAG	174
Qy	121	GGCGTCTCGGAGACTACCCGGGTGTCAAATCTTACTCTGGATATGTGGACACCTCTCC	180
Db	175	GGCGTCTCGGAGACTACCCGGGTGTCAAATCTTACTCTGGATATGTGGACACCTCTCC	234
Qy	181	GAGTCCCAATACCTTCTTGTGTTCTTCCGAGCCAGACATACCCAGAACCTGCACCTATC	240
Db	235	GAGTCCCAATACCTTCTTGTGTTCTTCCGAGCCAGACATACCCAGAACCTGCACCTATC	294
Qy	241	ACATTGTGGTTGAATGGTGGCCCTGGAGCGAATCTTTGTATGCGTCTCTTCGAAGAGTTG	300
Db	295	ACATTGTGGTTGAATGGTGGCCCTGGAGCGAATCTTTGTATGCGTCTCTTCGAAGAGTTG	354
Qy	301	GGCCCTTGCCATGTCAATTGCACTTTTGATGACTACATCAACCTCACTCGTGAAGAG	360
Db	355	GGCCCTTGCCATGTCAATTGCACTTTTGATGACTACATCAACCTCACTCGTGAAGAG	414
Qy	361	GCTCCAAATTTACTATTTCCTGCTCCAGCCATGGGAGTCGGCTTTTCATATAGTATAG	420
Db	415	GCTCCAAATTTACTATTTCCTGCTCCAGCCATGGGAGTCGGCTTTTCATATAGTATAG	474
Qy	421	GTTGATGGGTCCCAATTAACCTGTAACCTGCGGAGTCGAAATTCGAGCTTTGAGGAGTT	480
Db	475	GTTGATGGGTCCCAATTAACCTGTAACCTGCGGAGTCGAAATTCGAGCTTTGAGGAGTT	534
Qy	481	CAGGCGCGGTACCCCAACCAATGTATGCGACTCTGTATCGATACCTACCAATCTTGC	540

```
Db 535 CAGGCGGATACCAACCATTTGATGCCACTCIGATCGATATACACCAATCTTGGCGCAG 594
QY 541 GCAGCTTGGAGATCCCTGCAAGATTCCTTAGTGGACTACCTAGCTTGGACTTAGGGTG 600
Db 595 GCAGCTTGGAGATCCCTGCAAGATTCCTTAGTGGACTAGCTAGCTTGGACTTAGGGTG 654
QY 601 CAGTCTAAGGACTTTCAGTCTATGAGCGGAGAGCTATGAGGGCACTATGCTCCGCAATC 660
Db 655 CAGTCTAAGGACTTTCAGTCTATGAGCGGAGAGCTATGAGGGCACTATGCTCCGCAATC 714
QY 661 TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTAAATGTTGTTCA 720
Db 715 TTCAATCATTTTACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTAAATGTTGTTCA 774
QY 721 CTATATTTCACTCTCTGGGAATTTAATACGGCATCATCGAGGGGATCCAGGCCCT 780
Db 775 CTATATTTCACTCTCTGGGAATTTAATACGGCATCATCGAGGGGATCCAGGCCCT 834
QY 781 TACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACGGTC 840
Db 835 TACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGACGGTC 894
QY 841 TACAACATACATGAAGTTTGCAACCAAAATGATGCTGCCAGGATTTGATTTCCACC 900
Db 895 TACAACATACATGAAGTTTGCAACCAAAATGATGCTGCCAGGATTTGATTTCCACC 954
QY 901 TGCAACAGACAAACCCGACCGCATTCAGTACCTACGCCCTCGCGGAGCCACCAAC 960
Db 955 TGCAACAGACAAACCCGACCGCATTCAGTACCGCCCTCGCGGAGCCACCAAC 1014
QY 961 ATGTGCGAGGACAATTTGAGGGGCCATPACTACGCCCTTGTGCTGCTGCTGTATGAT 1020
Db 1015 ATGTGCGAGGACAATTTGAGGGGCCATPACTACGCCCTTGTGCTGCTGCTGTATGAT 1074
QY 1021 ATTGCGCATCATATGATGACCGGACTCCGCCAAGTTATTAACAATAATTTCTGGCAAG 1080
Db 1075 ATTGCGCATCATATGATGACCGGACTCCGCCAAGTTATTAACAATAATTTCTGGCAAG 1134
QY 1081 GACTCTGTCATGAGCGCTATCGGCTCAACATCACTACCGCTCAACAAATTAAGCGTC 1140
Db 1135 GACTCTGTCATGAGCGCTATCGGCTCAACATCACTACCGCTCAACAAATTAAGCGTC 1194
QY 1141 TACTACGCTTTCCAGCAACAGGCGACTTTGCTGCGCCAACTTCATCGAAGACCTCGAG 1200
Db 1195 TACTACGCTTTCCAGCAACAGGCGACTTTGCTGCGCCAACTTCATCGAAGACCTCGAG 1254
QY 1201 GAGATCTTGTCTCTCCGCTGGTGTCTCCCTCATCTATGCGGAGCGCGATACATCTGC 1260
Db 1255 GAGATCTTGTCTCTCCGCTGGTGTCTCCCTCATCTATGCGGAGCGCGATACATCTGC 1314
QY 1261 AACTGGTTGCGGGTCAGGCGCTTTCCTCGCTGGAATCTACTCCCAAGCCCGCATTC 1320
Db 1315 AACTGGTTGCGGGTCAGGCGCTTTCCTCGCTGGAATCTACTCCCAAGCCCGCATTC 1374
QY 1321 CGAAGCGAGGATACACGCGCTGAAAGTCAACGCGCTCGAGTATGGGAACCTCGGAG 1380
Db 1375 CGAAGCGAGGATACACGCGCTGAAAGTCAACGCGCTCGAGTATGGGAACCTCGGAG 1434
QY 1381 TATGTAATTTCTCTCTCCTGCTATGAGGAGGCGCATGAGTCCCATACCTACACAG 1440
Db 1435 TATGTAATTTCTCTCCTGCTATGAGGAGGCGCATGAGTCCCATACCTACACAG 1494
QY 1441 CCATCGCTCTCTCAATTTGTTTAAACGGACTATCTTCGGTTGGATATCGCAGAGGC 1500
Db 1495 CCATCGCTCTCTCAATTTGTTTAAACGGACTATCTTCGGTTGGATATCGCAGAGGC 1554
QY 1501 CAGAAGAGATCTGCGCCAGCTACAGAGCAATGGAAGCGGTACAGCTACGATACACAG 1560
Db 1555 CAGAAGAGATCTGCGCCAGCTACAGAGCAATGGAAGCGGTACAGCTACGATACACAG 1614
QY 1561 TCGTCGGTCCGCTGCTACGGCTACACCATGCTCCAGTGTGGTATG 1608
```

```
Db 1615 TCGTCGCTGCGCTGCGCTACGCTACGCTACGAGATGCTCCAGTGTGGTATG 1662
RESULT 2
AF394242 2245 bp DNA linear PLN 24-JUL-2001
LOCUS Aspergillus oryzae strain TK3 carboxypeptidase SI (cpl) gene,
DEFINITION complete cds.
ACCESSION AF394242
VERSION AF394242.1 GI:15004615
KEYWORDS Aspergillus oryzae
SOURCE Aspergillus oryzae
ORGANISM Aspergillus oryzae
REFERENCE 1 (bases 1 to 2245)
AUTHORS van den Broek, P.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,
Lausanne CH-1000, Switzerland
FEATURES
Location/Qualifiers
1..2245
/organism="Aspergillus oryzae"
/mol_type="genomic DNA"
/strain="TK3"
/db_xref="taxon:5062"
<1..>2245
/gene="cpl"
/product="serine carboxypeptidase 1"
join(1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)
/gene="cpl"
/EC_number="3.4.16.6"
/codon_start=1
/product="carboxypeptidase SI"
/protein_id="AAK77166.1"
/db_xref="GI:15004616"
/translation="MRGEFLSVLPVAASWALPGSTPASVGRRLPKNPTGVKLTIT
ANNVIRKEPGAEGVCTPGVSKSYGVDPSPESHFPFEPFARHPETAPITLWL
NGPGSDSLIGLFEELGCHVNSTEDITINHSNVEYNLLFSLPGLGSPYNDYD
GSINPTGVWNSPAGVGRYPTIDATIDITINLAARAAWELLOGFLGFLSDSRV
QSKDFLESTESYGHYAFENHFEYQNERANGVQNLNGLINGLIDRAIQ
APYDFEAVNTYGIKAVNETVYNTKFNANPNCODLITCKOTNRTALADALCA
EATNCRDNVEGYPVAFAGVYDIRHPYDDPTPSPYXKFLAKSVMDAIGVNTY
QSNNDVYVAFQGTGDFVNPFLDLEELALPVYSLLYGDADYPCNKGGAASLAA
NLSQAQRKSAGYTELKNGVYGETRHYGNFRTYVEAGHEVYYPYIASLQLENR
TIFGWDIAEGQKKIWPESYKNTGTATATHTQSSVPLPTATSSSVGMA"
1..54
sig_peptide
/gene="cpl"
```

```
BASE COUNT 548 a 603 c 524 g 570 t
ORIGIN
```

```
Query Match 33.0%; Score 531; DB 8; Length 2245;
Best Local Similarity 100.0%; Pred. No. 9.2e-301;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1078 AAGGACTCTGTCATGACGCTATCGCGTCAACATCACTACACCCAGTCCCAATATGAC 1137
Db 1709 AAGGACTCTGTCATGACGCTATCGCGTCAACATCACTACACCCAGTCCCAATATGAC 1768
QY 1138 GTCTACTAGCTTTCAGCAAAACAGGCGACTTGTGTGSCCAACTTCAICGAAGACCTC 1197
Db 1769 GTCTACTAGCTTTCAGCAAAACAGGCGACTTGTGTGSCCAACTTCAICGAAGACCTC 1828
QY 1198 GAGGAGATCCTTGCTCTCCCGTGGTGTCTCCCTCATCTATGCGGACCCGATACATC 1257
Db 1829 GAGGAGATCCTTGCTCTCCCGTGGTGTCTCCCTCATCTATGCGGACCCGATACATC 1888
QY 1258 TGCAACTGCTTGGCGGTCAGGCGCTTCCCTCGCTGCGAAGTACTTCCCAAGCGGCCAG 1317
```

Db 1889 TGCAACTGGTTGCGCGGTCTAGGCGGTTTCCTCGCTCGGAACACTACCCCAAGCGGCCAG 1948
QY 1318 TTCCGAAGCGCAGGGGTACAGCCCTGAAGTCAACGGCTCGAGTAIGGGGAAACTCGC 1377
Db 1949 TTCCGAAGCGCAGGGGTACAGCCCTGAAAAGTCAACGGCTCGAGTAIGGGGAAACTCGC 2008
QY 1378 GAGTATGGTAATTTCCTTCACTCGCGTCTATGAGGCAGGCCATGAAGTCCCATACTAC 1437
Db 2009 GAGTATGGTAATTTCCTTCACTCGCGTCTATGAGGCAGGCCATGAAGTCCCATACTAC 2068
QY 1438 CAGCCCATCGCCTCCCTGCAATTGTTAACGGACTATCTCGGTGGGATATCGCAGAG 1497
Db 2069 CAGCCCATCGCCTCCCTGCAATTGTTAACGGACTATCTCGGTGGGATATCGCAGAG 2128
QY 1498 GGCAGAGAAGATCTGGCCAGCTACAGACGANTGGAACGGCTACAGCTACGCATACA 1557
Db 2129 GGCAGAGAAGATCTGGCCAGCTACAGACGANTGGAACGGCTACAGCTACGCATACA 2188
QY 1558 CAGTCTCCGTGCGGTGCTACGGGTACCGCATGTCCAGTGTGGTATG 1608
Db 2189 CAGTCTCCGTGCGGTGCTACGGGTACCGCATGTCCAGTGTGGTATG 2239

Search completed: September 16, 2003, 21:25:04
Job time : 7741 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:06:30 ; Search time 148 Seconds
(without alignments)
4795.568 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 1608
Sequence: 1 cttccaggaagtacacggc.....gcattgtccagtgttggtatg 1608

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCITUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1608	100.0	1662	3	US-08-943-714-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-943-714-1
; Sequence 1, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION:
; APPLICANT: Blinkovsky, Alexander
; APPLICANT: Berka, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Dambmann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 61875780 No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA

ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,714
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4990.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1662
OTHER INFORMATION:
US-08-943-714-1

Query Match 100.0%; Score 1608; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CTTCCAGGAAGTACACCGCGCTCCGCTGGGTAGGAAGACAGCTACCCCAAGAACCCACCGGG	60
DB	55	CTTCCAGGAAGTACACCGCGCTCCGCTGGGTAGGAAGACAGCTACCCCAAGAACCCACCGGG	114
QY	61	GTCAGAGCTTTACAAACCGCAACAAATGTACCATCCGGTACAGGAACCCGGGGCAGAG	120
DB	115	GTCAGAGCTTTACAAACCGCAACAAATGTACCATCCGGTACAGGAACCCGGGGCAGAG	174
QY	121	GGCGTCGGCAGACTACCCCGGGTGTCAATCTACTCTGGATATGTCACACCTCTCCC	180
DB	175	GGCGTCGGCAGACTACCCCGGGTGTCAATCTACTCTGGATATGTCACACCTCTCCC	234
QY	181	GAGTCCCATACCTTCTTCTGGTTCTTTCGAAGCCAGACATAACCCAGAAATGACCTATC	240
DB	235	GAGTCCCATACCTTCTTCTGGTTCTTTCGAAGCCAGACATAACCCAGAAATGACCTATC	294
QY	241	ACATTGTGGTTGAATCGTGGCCCTCGAAGGATTCCTTGTATCGTCTCTTCGGAAGAGTTG	300
DB	295	ACATTGTGGTTGAATCGTGGCCCTCGAAGGATTCCTTGTATCGTCTCTTCGGAAGAGTTG	354
QY	301	GGCCCTTGCATGTCATTCGACTTTTGTAGTACATCAACCTCACCTCGTGGGAACGAG	360
DB	355	GGCCCTTGCATGTCATTCGACTTTTGTAGTACATCAACCTCACCTCGTGGGAACGAG	414
QY	361	GTCTCCAAITTTACTATTCCTTGCCAGCCATTCGGAGTGGGCTTTTCATATAGTATACG	420
DB	415	GTCTCCAAITTTACTATTCCTTGCCAGCCATTCGGAGTGGGCTTTTCATATAGTATACG	474
QY	421	GTGATGGGTCCATTAACTGAACTGGGTGCTCGAATAATTCGAGCTTTTCAGAGATT	480
DB	475	GTGATGGGTCCATTAACTGAACTGGGTGCTCGAATAATTCGAGCTTTTCAGAGATT	534
QY	481	CAGGGCCGGTACCAACCAATTTGATGCACCTCTCATGATACCTACCAATCTTCGCGAG	540
DB	535	CAGGGCCGGTACCAACCAATTTGATGCACCTCTCATGATACCTACCAATCTTCGCGAG	594
QY	541	GCGGCTTGGGAGATCCITGAAGGATTCCTTAGTGGACTACCTAGCTTGGACTAGGGTG	600

Db 595 GCGGCTTGGAGATCTCGAAGATTCCTTAGTGGACTACCTAGCTGGACTAGGGTG 654
QY 601 CAGTCTAAGGACTTCAGTCTATGAGCGGAGAGGCTATGGAGGCACTATGTCCTGCAATC 660
Db 655 CAGTCTAAGGACTTCAGTCTATGAGCGGAGAGGCTATGGAGGCACTATGTCCTGCAATC 714
QY 661 TTCAAATCATTTTACGAGCAGAAATGAGAGAAATGCCAAGGTAAGTGTAAATGGTGTTCAG 720
Db 715 TTCAAATCATTTTACGAGCAGAAATGAGAGAAATGCCAAGGTAAGTGTAAATGGTGTTCAG 774
QY 721 CTTAATTTCAACTCTCTGGGAATTTATTAAGCGCATCATCCAGAGCGGATCCAGGCCCT 780
Db 775 CTTAATTTCAACTCTCTGGGAATTTATTAAGCGCATCATCCAGAGCGGATCCAGGCCCT 834
QY 781 TACTACCCCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGAGCGTC 840
Db 835 TACTACCCCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGTCAACGAGAGCGTC 894
QY 841 TACAACCTACATGAAGTTTGGCAACCAATGCCAAATGGTTGCCAGGATTTGATTCAC 900
Db 895 TACAACCTACATGAAGTTTGGCAACCAATGCCAAATGGTTGCCAGGATTTGATTCAC 954
QY 901 TGC AAAACAGACAAACCGCAGCATAGCTGACTAGCTAGCCCTCTGCGCCGAGCCACCAAC 960
Db 955 TGC AAAACAGACAAACCGCAGCATAGCTGACTAGCTAGCCCTCTGCGCCGAGCCACCAAC 1014
QY 961 ATGTGAGGAGACAATTTGAGGGGCCATATAGCCCTTTGCTGGTGGTGTGTATGAT 1020
Db 1015 ATGTGAGGAGACAATTTGAGGGGCCATATAGCCCTTTGCTGGTGGTGTGTATGAT 1074
QY 1021 ATTCGGCATCCATATGATGACCGGACTCCGCCAAGTTATTACAACAAATTTCTGGCAAG 1080
Db 1075 ATTCGGCATCCATATGATGACCGGACTCCGCCAAGTTATTACAACAAATTTCTGGCAAG 1134
QY 1081 GACTCTGTATGAGGAGCTATCGGCTCAACATCACTACACCACTCCCAATATATGACGTC 1140
Db 1135 GACTCTGTATGAGGAGCTATCGGCTCAACATCACTACACCACTCCCAATATATGACGTC 1194
QY 1141 TACTAGGCTTCCAGCAACAGCGGACTTTGCTGGCCCACTTCATCGAAGACCTCGAG 1200
Db 1195 TACTAGGCTTCCAGCAACAGCGGACTTTGCTGGCCCACTTCATCGAAGACCTCGAG 1254
QY 1201 GAGATCCTTGCTCTCCCGCTGGGTGTCCTCCCTCACTATGCGGACGCCGATACATCTGC 1260
Db 1255 GAGATCCTTGCTCTCCCGCTGGGTGTCCTCCCTCACTATGCGGACGCCGATACATCTGC 1314
QY 1261 AACTGGTTCGGGGTTCAGCGCTTTCCCTCGCTCGCACTACTCCCAAGCGCCCACTTC 1320
Db 1315 AACTGGTTCGGGGTTCAGCGCTTTCCCTCGCTCGCACTACTCCCAAGCGCCCACTTC 1374
QY 1321 CGAAGCGCAGGCTACACGCCCTCGAAAGTCAACGGGCTGAGTATGGGAAACTCGCGAG 1380
Db 1375 CGAAGCGCAGGCTACACGCCCTCGAAAGTCAACGGGCTGAGTATGGGAAACTCGCGAG 1434
QY 1381 TATGTAATTTCTCCTTCACTCGGCTATGAGGAGGCCCATGAGTCCCATACACAG 1440
Db 1435 TATGTAATTTCTCCTTCACTCGGCTATGAGGAGGCCCATGAGTCCCATACACAG 1494
QY 1441 CCCATCGCTCCCTGCAATTTTAAACGGGACTATCTTCGGTTGGGATATCGCAGAGGC 1500
Db 1495 CCCATCGCTCCCTGCAATTTTAAACGGGACTATCTTCGGTTGGGATATCGCAGAGGC 1554
QY 1501 CAGAAGAGATGTGCCCGAGCTACAAGACGAATGGAACGGCTACAGCTACGCATACACAG 1560
Db 1555 CAGAAGAGATGTGCCCGAGCTACAAGACGAATGGAACGGCTACAGCTACGCATACACAG 1614
QY 1561 TCGTCGGTCCGCTGCTACGCTACCGCTACCGATCTCCAGTGTGGTATG 1608
Db 1615 TCGTCGGTCCGCTGCTACGCTACCGCTACCGATCTCCAGTGTGGTATG 1662

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:16:01 : Search time 517 Seconds
(without alignments)
7650.964 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 1608
Sequence: 1 cttccaggaagtacaccgcc.....gcatgtccagtgttggtatg 1608

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1660708 seqs, 1229959015 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description

No matches found

Search completed: September 16, 2003, 22:46:25
Job time : 517 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Search completed: September 16, 2003, 22:34:58
Job time : 4184 secs

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:04:35 ; Search time 4184 Seconds
(without alignments)
9340.726 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662
Perfect score: 1608
Sequence: 1 cttccaggaagatcacccgc.....gcattgcagtggtgtatg 1608

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152236056 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pla.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

No matches found

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 16, 2003, 14:42:53 ; Search time 7740 Seconds
(without alignments)
8499.054 Million cell updates/sec
Title: US-09-712-338-1_COPY_55_1662
Perfect score: 1608
Sequence: 1 cttcaggaaagtacaccgc.....gcattgcaggttggtatg 1608
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank:

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sy.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_mus.*
- 34: em_hgt_pln.*
- 35: em_hgt_rod.*
- 36: em_hgt_mam.*
- 37: em_hgt_vrt.*
- 38: em_sy.*
- 39: em_hgtgo_hum.*
- 40: em_hgtgo_mus.*
- 41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1608	100.0	1662	6	ARI29928	ARI29928 Sequence
2	937.8	58.3	2245	8	AF394242	AF394242 Aspergill
3	874.8	54.4	1656	6	AX534871	AX534871 Sequence
4	313.6	19.5	3150	6	AX534814	AX534814 Sequence
5	134.4	8.4	1872	6	AX534872	AX534872 Sequence
6	118.6	7.4	1665	6	AX534866	AX534866 Sequence
7	116.8	7.3	1581	6	AX534824	AX534824 Sequence
8	116.8	7.3	2940	6	AX534767	AX534767 Sequence
9	114.6	7.1	3080	6	AX534809	AX534809 Sequence
10	107.8	6.7	3221	6	AX534815	AX534815 Sequence
11	86.4	5.4	2503	6	A75535	A75535 Sequence 2
12	86.4	5.4	2503	6	I70282	I70282 Sequence 3
13	84	5.2	2002	6	I33983	I33983 Sequence 3
14	84	5.2	2002	6	I74375	I74375 Sequence 3
15	84	5.2	2002	6	I77239	I77239 Sequence 3
16	82.4	5.1	1611	6	AX534826	AX534826 Sequence
17	78.2	4.9	2660	6	AX534769	AX534769 Sequence
18	76	4.7	2068	6	I33982	I33982 Sequence 1
19	76	4.7	2068	6	I74374	I74374 Sequence 1
20	76	4.7	2068	6	I77238	I77238 Sequence 1
21	74.6	4.6	3850	6	E12103	E12103 DNA encodin
22	74.6	4.6	3850	8	PPRC1GEN	X87987 P.pastoris
23	72.4	4.5	660	8	CNS01BTO	AL114676 Botrytis
24	70.6	4.4	2214	8	PA067174	U67174 Pichia angu
25	68.4	4.3	1368	6	AX534850	AX534850 Sequence
26	68.4	4.3	1653	6	AX536918	AX536918 Sequence
27	68.4	4.3	2441	6	AX534793	AX534793 Sequence
28	68.2	4.2	4308	8	D86560	D86560 Schizosacch
29	68.2	4.2	37000	8	SPAC19G12	Z97209 S.pombe chr
30	67.6	4.2	1985	8	YSACARPEPY	M95182 Candida alb
31	66.2	4.1	2632	8	ARI29959	ARI29959 Sequence
32	66.2	4.1	2632	8	YSCPRCCPY	M15482 Yeast vacuo
33	66.2	4.1	15581	8	SC8175	X80836 S.cerevisia
34	65.6	4.1	2280	8	AB051820	AB051820 Aspergill
35	65.6	4.1	2509	8	AF085063	AF085063 Pichia an
36	63.2	3.9	1446	6	AX534852	AX534852 Sequence
37	63.2	3.9	2520	6	AX534795	AX534795 Sequence
38	61.8	3.8	1527	6	AX594884	AX594884 Sequence
39	61.8	3.8	2016	8	SCYBR139W	Z36008 S.cerevisia
40	61.8	3.8	2027	6	AX536454	AX536454 Sequence
41	61.8	3.8	29686	8	SC29711	X75891 S.cerevisia
42	61.8	3.8	50277	2	ACL38524_5	Continuation (6 of
43	59.6	3.7	1580	8	HVACXP112	X78878 Hordeum vul
44	59.4	3.7	4760	3	MOSAAGR	L46594 Aedes aegypt
45	59.2	3.7	1664	3	AK115283	AK115283 Ciona int

ALIGNMENTS

RESULT 1	ARI29928	1662 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	Sequence 1 from patent US 6187578.				
DEFINITION	ARI29928				
ACCESSION	ARI29928.1	GI:14117825			
VERSION					
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1662)				
AUTHORS	Blinkovsky,A., Berka,R., Rey,M., Golightly,E., Klotz,A., Mathisen,T.Erik., Dambmann,C. and Brown,K.M.				
TITLE	Carboxypeptidases and nucleic acids encoding the same				
JOURNAL	Patent: US 6187578-A 1 13-FEB-2001;				

FEATURES		Location/Qualifiers	
source	1..1662	/organism="unknown"	
BASE COUNT	396 a 468 c 398 g 400 t		
ORIGIN			
Query Match	100.0%;	Score 1608;	DB 6; Length 1662;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 1608;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	CTTCCAGAGTACACCGCGTCCGTCGGTGAAGACAGCTACCCAGAACCCACCGGG	60
Db	55	CTTCCAGAGTACACCGCGTCCGTCGGTGAAGACAGCTACCCAGAACCCACCGGG	114
QY	61	GTCAAGACTCTTACAACCGCAACAATGTCAACATCGGTACAGGAACCCGGGACAG	120
Db	115	GTCAAGACTCTTACAACCGCAACAATGTCAACATCGGTACAGGAACCCGGGACAG	174
QY	121	GGCGTCTGCAGACTACCCCGGGTGTCAAATCTACTCTGGATATGTCGACACCTCTCC	180
Db	175	GGCGTCTGCAGACTACCCCGGGTGTCAAATCTACTCTGGATATGTCGACACCTCTCC	234
QY	181	GAGTCCCATACCTTCTTCTGGTTCTTCCGAGGCGATTTCTTGATCGGTCTCTTCGAGAGTTG	240
Db	235	GAGTCCCATACCTTCTTCTGGTTCTTCCGAGGCGATTTCTTGATCGGTCTCTTCGAGAGTTG	294
QY	241	ACATTGTGTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTTCGAGAGTTG	300
Db	295	ACATTGTGTTGAATGGTGGCCCTGGAAGCGATTCTTTGATCGGTCTCTTCGAGAGTTG	354
QY	301	GGCCCTTGCATGTCATTCGACTTTTGAAGACTACATCAACCCCTCACTCTGGAACGAG	360
Db	355	GGCCCTTGCATGTCATTCGACTTTTGAAGACTACATCAACCCCTCACTCTGGAACGAG	414
QY	361	GTCTCCAAATTTACTATTCCTGTCCACGCCATTTGGAGTCGGTCTTTCATATAGTATACG	420
Db	415	GTCTCCAAATTTACTATTCCTGTCCACGCCATTTGGAGTCGGTCTTTCATATAGTATACG	474
QY	421	GTTCATGGGTCCATTAACCTGTAACTTGGGTGTCGAAAATTCGAGCTTTGAGGAGTT	480
Db	475	GTTCATGGGTCCATTAACCTGTAACTTGGGTGTCGAAAATTCGAGCTTTGAGGAGTT	534
QY	481	CAGGGCCGTATCCCAACCATATGACCACTCTGATCGATACTACCAATCTTGGCGACAG	540
Db	535	CAGGGCCGTATCCCAACCATATGACCACTCTGATCGATACTACCAATCTTGGCGACAG	594
QY	541	GCGCCTTGGAGATCCTGCAAGGATTCCTTAGTGAGTACCTAGCTTGGACTCTAGGGTG	600
Db	595	GCGCCTTGGAGATCCTGCAAGGATTCCTTAGTGAGTACCTAGCTTGGACTCTAGGGTG	654
QY	601	CAGTCTAAGGACTTCAGTCTATGACGCGAGAGCTATGAGGCGCACTATGTCCTGCAATC	660
Db	655	CAGTCTAAGGACTTCAGTCTATGACGCGAGAGCTATGAGGCGCACTATGTCCTGCAATC	714
QY	661	TTCATCATTTTACGAGACAAAGACAGAAATGCCAAGCGTAGTGTAAATGGTGTTCAG	720
Db	715	TTCATCATTTTACGAGACAAATGAGAGAAATGCCAAGCGTAGTGTAAATGGTGTTCAG	774
QY	721	CTTAATTTCAACTCTCTGGGAATTTAAACGGCAATCATCGACGAGGCGATCCAGGCCCT	780
Db	775	CTTAATTTCAACTCTCTGGGAATTTAAACGGCAATCATCGACGAGGCGATCCAGGCCCT	834
QY	781	TACTACCTGATTCGCTGTGAACAATACCTACGCTATCAAGGCTGTCAACGAGACCGTC	840
Db	835	TACTACCTGATTCGCTGTGAACAATACCTACGCTATCAAGGCTGTCAACGAGACCGTC	894
QY	841	TACAACCTACATGAATTTGCCAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACC	900
Db	895	TACAACCTACATGAATTTGCCAACCAATGCCAAATGGTTGCCAGGATTTGATTTCCACC	954
QY	901	TGCAAAACAGACAAACCGCACCGCATTTAGCTGACTATGCTGCGCCCTCTGCGCGGAGCCACCAAC	960
Db	955	TGCAAAACAGACAAACCGCACCGCATTTAGCTGACTATGCTGCGCCCTCTGCGCGGAGCCACCAAC	1014

961	ATGTGAGGACAATGTTGAGGGCCATACTACGCTTTGCTGGTCTGGTGTGTAATGAT	1020	
Db	1015	ATGTGAGGACAATGTTGAGGGCCATACTACGCTTTGCTGGTCTGGTGTGTAATGAT	1074
QY	1021	ATTTCGCATCCATATGATGACCCGACTCCGCCAAGTTATTACAACAATTTCTGCGAAAG	1080
Db	1075	ATTTCGCATCCATATGATGACCCGACTCCGCCAAGTTATTACAACAATTTCTGCGAAAG	1134
QY	1081	GACTCTGTATGAGGAGCTATCGCGCTCAACATCACTACACCGAGTCCCAATATGACGTC	1140
Db	1135	GACTCTGTATGAGGAGCTATCGCGCTCAACATCACTACACCGAGTCCCAATATGACGTC	1194
QY	1141	TACTACGCTTTCCAGCAACAGGCGACTTTGTCTGCCCAACTTCATCGAAGACCTCCGAG	1200
Db	1195	TACTACGCTTTCCAGCAACAGGCGACTTTGTCTGCCCAACTTCATCGAAGACCTCCGAG	1254
QY	1201	GAGATCCTTGTCTCCCGTGGTGTCTCCCTCATCTATGCGAGCGCGATTAACATCTGC	1260
Db	1255	GAGATCCTTGTCTCCCGTGGTGTCTCCCTCATCTATGCGAGCGCGATTAACATCTGC	1314
QY	1261	AACTGGTTCGGCGGTCAAGCGCTTTTCCTCGCTGCGAATCTCCCAAGCGCCGAGTTC	1320
Db	1315	AACTGGTTCGGCGGTCAAGCGCTTTTCCTCGCTGCGAATCTCCCAAGCGCCGAGTTC	1374
QY	1321	CGAAGCGGAGGTACACGCGCTGAAAGTCAACGGGCTCGAGTATGGGAAACTCGCGAG	1380
Db	1375	CGAAGCGGAGGTACACGCGCTGAAAGTCAACGGGCTCGAGTATGGGAAACTCGCGAG	1434
QY	1381	TATGGTAATTTCTCTCTCACTCGGCTATGAGGAGGCGCATGAAGTCCCATACTACCAG	1440
Db	1435	TATGGTAATTTCTCTCTCACTCGGCTATGAGGAGGCGCATGAAGTCCCATACTACCAG	1494
QY	1441	CCCATCGCTCCCTGCAATTTGTTAAACCGGACTATCTTCGGTTGGGATATCGCAGAGGC	1500
Db	1495	CCCATCGCTCCCTGCAATTTGTTAAACCGGACTATCTTCGGTTGGGATATCGCAGAGGC	1554
QY	1501	CAGAAGAAGATCTGGCCAGCTACAGAGCAATGGAACGGCTACAGCTACGCATACACAG	1560
Db	1555	CAGAAGAAGATCTGGCCAGCTACAGAGCAATGGAACGGCTACAGCTACGCATACACAG	1614
QY	1561	TCGTCGCTGGCGTCCCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACG	1608
Db	1615	TCGTCGCTGGCGTCCCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACG	1662

RESULT 2

AF394242

LOCUS

Aspergillus oryzae strain TK3 carboxypeptidase SI (cpl) gene,

complete cds.

ACCESSION

AF394242

VERSION

AF394242.1 GI:15004615

KEYWORDS

Aspergillus oryzae

SOURCE

Aspergillus oryzae

ORGANISM

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Trichomiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE

1 (bases 1 to 2245)

AUTHORS

van den Broek, P.

Direct Submission	
Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,	
Lausanne CH-1000, Switzerland	
Location/Qualifiers	
1..2245	
/organism="Aspergillus oryzae"	
/mol_type="genomic DNA"	
/strain="TK3"	
/db_xref="taxon:5062"	
<1..>2245	
/gene="cpi"	
Join(<1..349,409..513,576..692,743..866,941..956,	
1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)	


```

/gene="cpl"
/product="serine carboxypeptidase 1"
join(1..349,409..513,576..692,743..866,941..956,
1008..1124,1187..1237,1296..1449,1522..1580,1670..2245)
/gene="cpl"
/EC_number="3.4.16.6"
/codon_start=1
/product="carboxypeptidase s1"
/protein_id="AAK7166.1"
/db_xref="GI:15004616"
/translation="MRGYEFLSVLPLVAASWALPGSTPASVGRRLKPNKPTGVKLTIT
ANNVTRYKPGAGVCEITTPGVKSYGYVDIIPESHFFWFEEFARNPETAPITLWL
NGGPGSLLGLEEFELPGCHVNSTFDYINPHSWNEYSNLLFLSOPLGVSFYNQDVL
GSINPTGVVSVSFAVCGOKRPTIDATLIDTNLAAEAAWELIIGFLSLPLSDSV
QSKDFSLWTSYSGHGYPAFENHFYONERIANGSVNGVQLNENSLGILNIDEALQ
APVYPEAVNTYGIKAVNETVYNKPFANMENGCDLLSTCKQTKRIALADALQ
EATNMRDNVEGPIYAFAGGVYDIRHPYDDPTPPSYNKLPLAKDSYMDAIGVINNT
QSNVDYIAQQQGFDFPNFIEDLEILALPVRSIYGLADYICNMFQGVASLAA
NYSQAAQFRSAGYTPLVKNGVEXGETREYGNFSFTRVYEAGHEVPYQPIASLQLENR
TIFGWDIAEGKKIWPSTKNGTATATHTQSSVPLPTATSMSSVGM"
sig_peptide
1..54
/gene="cpl"
BASE COUNT 548 a 603 c 524 g 570 t
ORIGIN

Query Match 58.3%; Score 937.8; DB 8; Length 2245;
Best Local Similarity 73.5%; pred. No. 1e-231;
Matches 1606; Conservative 0; Mismatches 2; Indels 577; Gaps 9;

QY 1 CTTCCAGGAGTACACCGCGTCCGCTGGTAGAGACAGCTACCCAGAACCACCGGG 60
DB 55 CTTCCAGGAGTACACCGCGTCCGCTGGTAGAGACAGCTACCCAGAACCACCGGG 114
QY 61 GTCAAGACTCTTCAACCGCAACATGTCCACCATCGGTTACAGGAAACCGGGGAG 120
DB 115 GTCAAGACTCTTCAACCGCAACATGTCCACCATCGGTTACAGGAAACCGGGGAG 174
QY 121 GCGCTCTCGGAGACTACCGCGGTGTCAAATCTACTCTGGATATGTCGACACTCTCC 180
DB 175 GCGCTCTCGGAGACTACCGCGGTGTCAAATCTACTCTGGATATGTCGACACTCTCC 234
QY 181 GAGTCCCATACCTCTTCTGTTCTTTCGAAAGCAGACATAACCCAGAACTGACCTATC 240
DB 235 GAGTCCCATACCTCTTCTGTTCTTTCGAAAGCAGACATAACCCAGAACTGACCTATC 294
QY 241 ACATTGTGTTGAATGTGTCCTGGAAGCGATCTTGTGATCGTCTCTCGA----- 293
DB 295 ACATTGTGTTGAATGTGTCCTGGAAGCGATCTTGTGATCGTCTCTCGAAGGTTG 354
QY 294 -----AGAGTTGG 301
DB 355 CTTGACGCTGTTACAAGTATGCTCTTTTAGGTCGCGTTAACCTTTGTTAGATTGG 414
QY 302 GCGCTTGCATGCAATTCGACTTTGATGACATACATCAACCTCACTCTGGAAGAGG 361
DB 415 GCGCTTGCATGCAATTCGACTTTGATGACATACATCAACCTCACTCTGGAAGAGG 474
QY 362 TCTCCAAATTACTATTCTCTCCAGCCATTGGAGTC----- 399
DB 475 TCTCCAAATTACTATTCTCTCCAGCCATTGGAGTCGATGGAATGCGGACCCCTCC 534
QY 400 -----GGCTTTTCATATAGTATAC 419
DB 535 CCTACTGGGCTCTCCCATATTGACGAGGTGCTCCGCTAGGCTTTTCATATATGATAC 594
QY 420 GGTGTATGGTCAATTAACCTCTGAATGGGTGCTGGAATAATCGAGCTTTTCAGAGG 479
DB 595 GGTGTATGGTCAATTAACCTCTGAATGGGTGCTGGAATAATTCGAGCTTTTCAGAGG 654
QY 480 TCAGGCGCGGTACCCAAACCATTCATCCACTCTGATC----- 516
DB 655 TCAGGCGCGGTACCCAAACCATTCATCCACTCTGATCGGTATCGGTTCTCTCGGTCAC 714

```

```

QY 517 -----CATACIACCAATCTTGGCCGAGAGGCCCTTGG 549
DB 715 ATGCTTCTATTACTGATACGATAGCATATACCAATCTTGGCCGAGAGGCCCTTGG 774
QY 550 GAGATCCTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTCTAGGCTCAGTCTAAG 609
DB 775 GAGATCCTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTCTAGGCTCAGTCTAAG 834
QY 610 GACTTCAGTCTATGACGAGGAGCTATGGAG----- 640
DB 835 GACTTCAGTCTATGACGAGGAGCTATGAGGAGTGGTTCCTCTCTCGAAATCTTCA 894
QY 641 -----GGCACTATGGTCTCTG 655
DB 895 AACGGCCGTGATGTTGGGTGATAGGTTCTCTTTCGTTCGAAAGGAGCTATGGTCTG 954
QY 656 CA-----TTCATCA 664
DB 955 CAGTTCAGTCTACCTGGACGATGCTCAAGAACGACAGACTAAGAACGTAGTCTTCA 1014
QY 665 ATCAATTTTACGACGAGAAATGTCACACGCTAGTGTAAATGGTCTCAGCTTCA 724
DB 1015 ATCAATTTTACGACGAGAAATGTCACACGCTAGTGTAAATGGTCTCAGCTTCA 1074
QY 725 ATTTCAACTCTCTGGATATTAAACGGCATCATCGAGGCGATCC----- 772
DB 1075 ATTTCAACTCTCTGGATATTAAACGGCATCATCGAGGCGATCATCGAGGCGATCATCGG 1134
QY 773 -----AGGCCCTTCA 782
DB 1135 GAAACGCCCATAGAACATCTCCGACAAACGCTGCTAACGATGCCCATAGGCCCTTCA 1194
QY 783 CTACCTCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCT----- 825
DB 1195 CTACCTCTGAATTCGCTGTGAACAATACCTACGGTATCAAGGCTGAAGTTCCTCTTATG 1254
QY 826 -----GTCAACGAGACCGTCTACA 844
DB 1255 CGCATGAGATTATATCTCTGATATCTATCACCATTGAAAGGCTCAACGAGACCGTCTACA 1314
QY 845 ACTACATGAAGTTTGCACCAACAAATGTCGACGAGTTTGAATTTCCACCTGCA 904
DB 1315 ACTACATGAAGTTTGCACCAACAAATGTCGACGAGTTTGAATTTCCACCTGCA 1374
QY 905 AACAGACAAACCGCACCGCATTTAGCTGACTACGCGCTCTCGCGGAAAGCCACCAACATG 964
DB 1375 AACAGACAAACCGCACCGCATTTAGCTGACTACGCGCTCTCGCGGAAAGCCACCAACATG 1434
QY 965 GCAGGGACAATGTT----- 978
DB 1435 GCAGGGACAATGTTGGTACGTTCTTTCCCTTGGACACAAAGGTATATGATTCACAAA 1494
QY 979 -----GAGGGGCCATACCTAGCGCTTTGGCTTGGTGGTGGTGG 1012
DB 1495 TAGTGGGACAATAACACGAAATAAGAGGGCCATACCTAGCGCTTTGGCTTGGTGGTGG 1554
QY 1013 TGTATGATTTCCGATCCATATGAT----- 1038
DB 1555 TGTATGATTTCCGATCCATATGATGTAAGTAGAGCTTTTACITCTATTTTACACACCGC 1614
QY 1039 -----GACCC 1043
DB 1615 ATCAAAATGATGGTTCAATGTTGATCGAATAAAATCCCGCTAATATGTTATGAGACCC 1674
QY 1044 GACTCCGCCAAGTATTACAAATAATTTCTGGCAAGGACTCTGTGATGACGCTATCGG 1103
DB 1675 GACTCCGCCAAGTATTACAAATAATTTCTGGCAAGGACTCTGTGATGACGCTATCGG 1734
QY 1104 CGTCAACATCAACTACACCGAGTCCCAATAATAGCTTACTAGCGTTTCCAGCAACAGG 1163
DB 1735 CGTCAACATCAACTACACCGAGTCCCAATAATAGCTTACTAGCGTTTCCAGCAACAGG 1794
QY 1164 CGACTTTGTTGCGCCCAACTTCATCGAAGACCTTCGAGGAGATCCITGCTCTCCCGTGG 1223

```

```
Db 1795 CGACTTTGCTGGCCCACTTCATCGAAGACCTCGAGGAGATCCTTTGCTCTCCCGTGGC 1854
Qy 1224 TGTCTCCCACTATATGGGACGCCGATACATCTGCAACTGTTGCGGCGTCAAGCGGT 1283
Db 1855 TGTCTCCCTCATCTATGGGACGCCGATACATCTGCAACTGTTGCGGCGTCAAGCGGT 1914
Qy 1284 TTCCCTCGCTGGCACTACTCCCAAGCGCCCAAGTTCGGAAGCGCAGGTFACAGCCCT 1343
Db 1915 TTCCCTCGCTGGCACTACTCCCAAGCGCCCAAGTTCGGAAGCGCAGGTFACAGCCCT 1374
Qy 1344 GAAAGTCAACGGCGCTCGAGTATGGGAAACTCGCGAGTATGGTAAATTTCTCTCACTCG 1403
Db 1975 GAAAGTCAACGGCGCTCGAGTATGGGAAACTCGCGAGTATGGTAAATTTCTCTCACTCG 2034
Qy 1404 CGTCTATGAGCGAGCCATCAAGTCCCATCTACAGCCCATCGCCTCCCTGCAATGT 1463
Db 2035 CGTCTATGAGCGAGCCATCAAGTCCCATCTACAGCCCATCGCCTCCCTGCAATGT 2094
Qy 1464 TAACCGGACTATCTCGGTGGGATATCGAGAGGGCCAGAAAGATCTGGCCAGCTA 1523
Db 2095 TAACCGGACTATCTCGGTGGGATATCGAGAGGGCCAGAAAGATCTGGCCAGCTA 2154
Qy 1524 CAAGACGAATGAGCGCTACAGCTACGCATACACATACAGTCTCGTGGCGCTGCCTAGGC 1583
Db 2155 CAAGACGAATGAGCGCTACAGCTACGCATACACATACAGTCTCGTGGCGCTGCCTAGGC 2214
Qy 1584 TACCAGCATCTCCAGTGTGGTATG 1608
Db 2215 TACCAGCATCTCCAGTGTGGTATG 2239
```

RESULT 3

AX534871 LOCUS AX534871 1656 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 109 from Patent WO02068623.
ACCESSION AX534871
VERSION AX534871.1 GI:25261462

KEYWORDS

Aspergillus niger

SOURCE

Aspergillus niger

ORGANISM

Aspergillus niger

REFERENCE

1 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

AUTHORS

Edens, L., dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A.,

Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,

Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S.,

Kemmer, W., Tan, P., Stiebler, J. and Albang, R.

Novel genes encoding novel proteolytic enzymes

Patent: WO 02068623-A 109 06-SEP-2002;

DSM N.V. (NL)

FEATURES

Location/Qualifiers

1..1656

/organism="Aspergillus niger"

/mol_type="genomic DNA"

/db_xref="taxon:5061"

411 a 426 c 409 g 410 t

BASE COUNT

411 a

426 c

409 g

410 t

ORIGIN

Query Match 54.4%; Score 874.8; DB 6; Length 1656;
Best Local Similarity 71.7%; Pred. No. 2.1e-215;
Matches 1146; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

Qy 2 TTCCAGGAAGTACACCGCGCTCGGTGAGACAGCTACCCAAAGAACCCACCGGG 61
Db 56 TCCCGAGATGATGTGTCATCTACGATAGAGAGCGAGTTACCAAAAGCTCCACTGGC 115
Qy 62 TCAAGACTCTTACACCGCAACAAATGTCCACTCCGGTACAGGAACCCGGGCGAGG 121
Db 116 TCAATCGATAAAACCCCAACAAATGTCTATCAGTATATAAGAACCCAGAACCGAAG 175
Qy 122 CGGTCTGGAGACATACCCCGGGTGTCAAATCTACTCTGATATGTGACACCTCTCCG 181

```
Db 176 GAATTTGTGAGACAACACCTGGGGTCAAAATCAATACTCCGGAATATGTCTGATCTTTCCGCAG 235
Qy 182 AGTCCCATACCTCTCTGCTTCTGAGCCGACACATAACCCAGAAACTGCACCTATCA 241
Db 236 AGTCCGATACATCTTCTTTTGGTTTTCGAGTACGCGTACCCGGAATGATCCAGTGA 295
Qy 242 CATTTGGTTGAATGTTGGCCCTTGGAGACGATCTTTGATCGGTTCTTTCAGAGATTTGG 301
Db 296 CTCGTGGCTGAATGTTGGCCCTTGGAGACGATCTTCTGATTTGGGCTTTTGAAGAGTTGG 355
Qy 302 GCCCTGCCATGTCATCAATTCGACTTTTGATGACTACATCAACCTCTACCTGTGAACAGG 361
Db 356 GTCCGTGTACATCACACAGAGTACGAATCAATCAATCAATCAATCAATCAATCAATCAAT 415
Qy 362 TCTCCAAATTTACTATCTCTCCAGCCATTTGGGAGTCGGCTTTTTCATATAGTATACGG 421
Db 416 TCACCAATCTCTTTCTGTCTCAGCCCTCGGTGTGGGTCTCTTACAGTGAACCG 475
Qy 422 TTGATGGGTCCATTAACCTCTTAACCTGGGTGTGCGAAATTCGAGCTTTGACAGGATTC 481
Db 476 AGCGCGGTCTCTTGAATCCATTTACTGAGCGGTGCGAGAGCGCTCTCTTGTCTGGAGTTC 535
Qy 482 AGGCGCGTACCCCAACCATTTGATGCCACTCTGTATCGATACTACCAATCTTGCCTCCAGG 541
Db 536 AGGTCTGATACCCAGTTTATGATGCCACTATCATCGACACACCGATATGCTGCACCG 595
Qy 542 CCGCTTGGGAGATCTGCAAGGATTCCTTACTGAGCTACCTAGCTTGGACTCTAGGTC 601
Db 596 CAACCTGGAGGTCTCTCAGGGCTCTCAGTGGCTTCTCCAGCTTATGTCAGTAGATTCGGAAGTCA 655
Qy 602 AGTCTAAGGACTTCAGTCTATGAGCGAGACTATGAGGCGCACTATGGTCTGCTGATCT 661
Db 656 AGTCCAAAGGAGTTCAACCTGTGGACAGAGTATAGGAGGACACTATGGACAGCGTCT 715
Qy 662 TCATCATTTTACGACAGCAATGAGAAATGTCACAGGTAGTGTAAATGCTGTTCAGC 721
Db 716 TCAATCATTTCTACGAGCAAAATTCGAAGATCGGTAGCGGGAGTAGTCATGCGGCTCCAC 775
Qy 722 TTAATTTCACTCTCTGGGAATTAATACGGCATCATCGACGAGCGCATCCAGGCCCT 781
Db 776 TGAATTTAATCTCCCTCGGATTAATCAACGSCATCATTTGATGCGCGATTCAGCAGACT 835
Qy 782 ACTACCTGATTCGCTGTGAACAAATACCTACGATATCAAGGCTGTCAACAGACCGTCT 841
Db 836 ACTACGAGACTTTTCCGCTTAATATACATATGAATCAAAASCTGTCAATACACACTGT 895
Qy 842 ACAACTACATGAAGTTTGCACCAAAATGTCAGGATTTGCGAGGATTTGATTTCCACT 901
Db 896 ACAACTATATGAGTTTCCCAACACGATGCGCAATGATGCGAGATCAGGTTGCTGCT 955
Qy 902 GCAACACAGACAAACCGCACCGCATTTAGCTAGCTAGCGCTCTGCGCGAAGACCCACACA 961
Db 956 GTAAATTGACCAATAGGACCTCGCTTCTGATTTATGCTATATGATACAGACAGCCAATA 1015
Qy 962 TGTGAGGACAATTTGAGGGGCCATACCTACGCTTTGCTGGTGTGCTGTATGATA 1021
Db 1016 TGTGAGGACAATTTGAGGGGCCATACCTACGCTTTGCTGGTGTGCTGTATGATA 1075
Qy 1022 TGTGAGTCCATATGATGACCGCATCTCCGCAAGTTATTAACAACAAATTTCTGCGCAAGG 1081
Db 1076 TGTGAGCGCCCTCAATGATGACCGCCCGCTCTACTTTGTTGCTACCTCAAGAAG 1135
Qy 1082 ACTCTGTATGAGGCTATCGCGGTCAACATCAACTACACCGATTCCTCAATATGACGCT 1141
Db 1136 ACTCAGTCTATGATGATCGCGGTGACATTAATACACCGAGTCCAGCGCGCAAT 1195
Qy 1142 ACTAGCTTTCCAGCAACACAGCGCTTTGTCGCCCAACTTCATCGAAGACCTTCGAGG 1201
Db 1196 ATTATGCAATTCAGACAGCGCGACTTTGATGGCCGAATTTCAATGAGACCTTCGAG 1255
Qy 1202 AGATCTGTCTCTCCCGTGGTGTCTCCCTCATCTATGGGACCGCCGATTAATCTGCA 1261
Db 1256 AGATCTCCAACTCCCGGTACGCTGTCTGATCTACGGGATCGGACTATATCTGTA 1315
```

Qy	1262	ACTGGTTGGCGGTGAGCGCGTTTCCTCGCIGCGAACTACTCCACAGCCGCCAGTTCC	1321
Db	1316	ACTGGTTGGCGGTGAGCGCCATCTCACTCGAGTTAACTACCCCATGCAAGTCAGTTCC	1375
Qy	1322	GAAGCGCAGGATACACGCCCTGAAATCAACGGCGTCGAGTAGTGGGGAACACTCGCGAGT	1381
Db	1376	GTGCAGCGGGATACACACCCATGACACTAGATGGGTGCGAATACGGTGAAGTCGCGAGT	1435
Qy	1382	ATGGTAATTTCCTCTCACTCGCGGTCTATGAGCGAGGCCATGAAGTCCCATACTACACGC	1441
Db	1436	ATGGCAACTTTTCGTTTCACCGCGGTATATCAGGCTGGGCACGAGGTTCCATCTATCAAC	1495
Qy	1442	CCATCGCCTCCCTGTCAATTCGTTAAACGGGACTATCTTCGGTTGGGATATCGCAGAGGCC	1501
Db	1496	CGATCGCAGGTTGCAGCGCTGTTCAACGGTACTTTATTGTGGATGGGAATTTGCAAGCGGTA	1555
Qy	1502	AGAAGAAGATCTGGCCCCAGCTACAAGACGAATGGAACGGCTACAGCTTAACGATACACAGT	1561
Db	1556	CAACTCAGATTTGGCCCCGAATATAGCACCAACGGGACATCGCAGGCTACACACCGAGT	1615
Qy	1562	CGTCCGTGCGGTGCGCTTACGGCTACACAGCATGTCCAGT	1599
Db	1616	CGTTCGTGCCACTGTCCACCGCGGTTCGAGTACCGGTCAAT	1653

RESULT	4
LOCUS	AX534814 Sequence 52 from Patent WO02068623.
DEFINITION	AX534814
ACCESSION	AX534814.1 GI:25261343
VERSION	
KEYWORDS	
SOURCE	Aspergillus niger
ORGANISM	Aspergillus niger Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE	1
AUTHORS	Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A., Kimpel, E., Kludbauer, S., Wagner, C., Fritz, A., gustedt Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albang, R. Novel genes encoding novel proteolytic enzymes Patent: WO 02068623-A 52 06-SEP-2002;
TITLE	DSM N.V. (NL)
JOURNAL	Location/Qualifiers
FEATURES	1..3150 /organism="Aspergillus niger" /mol_type="genomic DNA" /db_xref="taxon:5061"
SOURCE	
BASE COUNT	801 a 778 c 738 g 833 t
ORIGIN	

	Query Match	19.5%;	Score 313.6;	DB 6;	Length 3150;
	Best Local Similarity	72.5%;	Pred. NO. 5.6e-70;		
	Matches 406;	Conservative 0;	Mismatches 154;	Indels 0;	Gaps 0;
QY	1037	ATGACCGGACTCGGCCAAGTATTACACAAATTTCTGGCAAGGACTCTGTCATCGACG	1096		
Db	2194	AGGACCCGACCCGCGCTCTACTTTGTGACTACCTCAAGAAAGACTCAGTCATGGATG	2253		
QY	1097	CTATCGGGCTCAACATCACTACACCCAGTCCCAATAATGACGTCTACTACGTTTCCAGC	1156		
Db	2254	CTATCGGCTGGACATTAACATACCCGAGTCCAGCGCGGAAGTATATATGCATTTCCAGC	2313		
QY	1157	AAACAGCGGACTTTGTTGGCCCAACTTCATCGAAGACCTCGAGGAGATCCCTTGCCTCC	1216		
Db	2314	AGACCGGCGACTTTGATGGCCGAATTTCAATTGAGGACCTCGAAGAGATCCTCCAACTCC	2373		
QY	1217	CGTGGCTGTCCTCCCTACTATGGGAGCGCGATACATCTGCACTGGTTCGGCGGTC	1276		
Db	2374	CGGTACCGGTGCTGTGATCTACGGCGATGCCGACTATATCTGTAACTGGTTCGGCGTC	2433		

QY	1277	AGGCCGTTTCCTCGCTGCGAACTACTCCCAAGCGGCCAGTTCGGAAGCGCAGGGTACA	1336
Db	2434	AGGCCATCTCACTCGCAGTTAACTAACCCCATGCAGCTCAGTTCGCTGCAGCGGATACA	2493
QY	1337	CGCCCTGGAAGTCAACGGCTCGAGTATGGGAAACTCGGGAGTATGGTAATTTCCT	1396
Db	2494	CACCATGACAGTAGATGGGTCCGAATACCGTGAGACTCGCGAGTATGGCAACTTTTCGT	2553
QY	1397	TCACTCGGCTCATGAGCGAGGCCATGAAGTCCCACTACACAGCCCATCGCTCCCTGC	1456
Db	2554	TCACCCGGTATATCAGCTGGGACAGAGTTCCAACTATCAACCATCGCAGCGTTGC	2613
QY	1457	AATTGTTTAAACCGGACTATCTTCGGTTGGATATCGCAGAGGGCCAGAAGAAATCTGGC	1516
Db	2614	AGCTGTTCAACCGTACTTTATTTGGATGGGATATTCGAGCGGTACAACTCAGATTGGC	2673
QY	1517	CCAGCTACAGACGAATGGACGCTACAGCTAGCAYACACACTCGCTCGGTGCGCGTGC	1576
Db	2674	CCGAATATAGCAACACGGGACATFCGAGGCTACACACAGGGAGTCTGTTGGTGCCACTGT	2733
QY	1577	CTACGGGTACACGCAATGCC	1596
Db	2734	CCACGGCGTCGAGTACCACC	2753
RESULT 5			
AX534872			
LOCUS	AX534872	1872 bp	DNA linear PAT 22-NOV-2002
DEFINITION	Sequence 110 from Patent WO02068623.		
ACCESSION	AX534872		
VERSION	AX534872.1 GI:25261464		
KEYWORDS	.		
SOURCE	Aspergillus niger		
ORGANISM	Aspergillus niger		
REFERENCE	1 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.		
AUTHORS	Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A., Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gusted Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albang, R.		
TITLE	Novel genes encoding novel proteolytic enzymes		
JOURNAL	Patent: WO 02068623-A 110 06-SEP-2002; DSM N.V. (NL)		
FEATURES	Location/Qualifiers		
source	i. 1872		
	/organism="Aspergillus niger"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:5061"		
BASE COUNT	402 a	434 c	570 g 466 t
ORIGIN			

Query Match	8.4%;	Score 134.4;	DB 6;	Length 1872;
Best Local Similarity	48.1%;	Pred. No. 1.4e-23;		
Matches	644;	Conservative 0;	Mismatches 661;	Indels 33; Gaps 8;
QY	186	CCATACCTTCCTTCGGTCTCTCGAAGCCGACAT-----AACCCAGAAACTGCACCTAT	239	
Db				
Db	282	CAATACCTTTTCGGTACTTTCCTCCGCCATACCACAACAATGATACATCCCCACT	341	
QY	240	CACATTGTGTTGAAATGGTGGCCCTGGAAGCGATTCTTTTGATCGGTCCTTCGAAAGATT	299	
Db	342	CACCATCTGGATGAACGGGGGGCGGGCGGATCCICCATGATGGGCTATTTCAAGAGAA	401	
QY	300	GGSCCCTTCGCATGTCAATTGCAGCTTTTGATGACTACATCAACCCACAC---TCGTGGAA	356	
Db	402	CGGGCCATGTACTGTCAATACGCACTCGAATTCACGGCCTATTAATCCCTGGTCGGGAA	461	
QY	357	CGAGGTCTCCAAATTTACTATTCCTGTCCGACCAATGGGAGTCGGTTTTCATATAGTGA	416	
Db	462	TGAGTAGTCGATAIGTTGTAATATTGACAGCCGGTGCACAGCGGATTTAGTATATGATCT	521	
QY	417	TACGGTTGATGGTCCATTAACCTCTAACTGGGTCGTGCGAAATATCGAGCTTTTCGAGG	476	

Db 522 GTTGAGGAATGGGACGTTAGAT---TTGAATGAGACGTTTTTTGGTGGGACGTTGGCG- 576
Qy
Db 477 AGTTACGGCGCGGTACCAACCAATGATGACCACTCTGATGCACTACCAATCTTGCGCG 536
Db 577 -AGTCAGATGTGATGGGCGGTGATGGGACGGTTAATGGGGAAGGCGCTTTGGGT 635
Qy 537 AGAGGCGGCTGGGAGATCTGCAAGGATTCCTTAGTGGACTACTAGCTTGGACTTAG 596
Db 636 TGGTTGACAGTTGGTTGGTGAATTCCTGAAATATGTTCTCTGTTGACGGGAATGG 695
Qy 597 GGTGACACTTAAGACCTTACGCTTATGGACGAGCTATGGAGGCACTATGTCCTGC 656
Db 696 TGGTGGTGATACAGGGTGAGTATATGACGGAGTATATGGGGACGGTATGACCGCG 755
Qy 657 ATCTTCAATCAATTTTACAGCAGATGAGAGAAATGGCAACGGTAGTGT---TAATGG 713
Db 756 ATACACGGCGCTCTTACAGAGATGAATGAGAGGATGAGAGTGGGGAGTAAACCGG 815
Qy 714 TGTTCAGCTTAATTTCAACTCTCTGGGAATATTAAACGGCATATGACGAGGCGATCCA 773
Db 816 GAAGAAGATCCATTTGATAGCGGCGGCAATTAATGAGTGGTGTGATTTACTCTGCA 875
Qy 774 GGCCCTTACTACCTGAATTCGCTGTGAACAATACCTACGCTATCAAGCTGTCAACGA 833
Db 876 GGTCCCTTCTCTGAGCAGCGGTATAACAATACGATATGGGATCGAGGGAATCAATCG 935
Qy 834 GACGGTCTACACTACATGAATTTGGCAACCAATGCCAAATGGTTGCCAGGATTTGAT 893
Db 936 CAGCTCTACGACGGGCTATGGATAGTTGAGCAAGCCTGGCGGTTGACGGGATATGAT 995
Qy 894 TTCCACCTGCAACAGACAAACCGACCGCATTAAGCTGACTTAACGCTCTGCGCGGAAGC 953
Db 996 CATGAGTGTGCGCATCTGCGGAGCTCGG---AGATCCCTCATCATCTGCGAGGAGC 1052
Qy 954 CACCAACATGTGACGAGCAATGTGAGGGGCCAATACCTACGCTTGTGCTGTGCTGPT 1013
Db 1053 GTGCGACTGTGTCGGGAGATCAAGAGCGCTGTATACGAATACCTCCGGCGGAGGATA 1112
Qy 1014 GTATGATATTCGATCCATATGATGACCGGACCTCCGCAAGTATACACAAATTTCT 1073
Db 1113 CTAGGACATAGCGATTTACGCGGATGACGCTCTGTCGCTTACTTCGTCGGGTCTT 1172
Qy 1074 GGCAAGAGCTCTGTATGACGCTATCGGCGTCAACATCAACTACACCCAGTCCCAATA 1133
Db 1173 GAATCGCCATGGGTGCAAGAGGCACTTGGGGTCCCGGTGAACATACCAATGCTGAGA 1232
Qy 1134 TGAGCTCTACTACGCTTCCAGCAACAGGCGACTTTGTCTGGCCCAACITTCATCGAGA 1193
Db 1233 GGCAAGTGGGAACAGTTTCGCTCGAGGGCGGATATCCGGGAAATGATCCCGGGAAT 1292
Qy 1194 CCTGAGGAGA-----TCCCTTGTCTCCCGTGGCTCTCCCTCATCTATGGGA 1244
Db 1293 GATCGGGGATATGGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1352
Qy 1245 CGCGGATATCACTGCAACTGGTTCGCGGCTGAGCGGCTTTCCCTCGCTGCGCACTATC 1304
Db 1353 CCGGACTATGCTTGTCCGTTGGCGCGGGGGAAGATGTCAGCGCTGCTGCTGCTGCTGCTG 1412
Qy 1305 CCAAGCGCCGAGTTCCGAGCGAGGATACACCGCCCTGAAAGTCAACGCGCTGAGTA 1364
Db 1413 GGATTCGGGAGAAGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1472
Qy 1365 TGGGG---AAACTCGGAGTATGTTAAATTTCTCTCACTCGGCTCTATGAGGCGGCA 1421
Db 1473 TGGGGTCTACTAGGAGATGAGGAATCTGCTTACCGGCTGCTTACGCGGGGCA 1532
Qy 1422 TGAAGTCCCACTACCAAGCCATCGCTCCCTGCAATGTTTAAACCGCACTATCTCGG 1481
Db 1533 TGAGGTGCCATTTTATCAGCCGGAACCGGCTATGAGATTTTAAATCGCTGCTGCTGCTG 1592
Qy 1482 TTGGGATATCCGAGGG 1499
|||||

Db 1593 TTGGATATTCGACGGG 1610
RESULT 6
AX534866
LOCUS AX534866 1665 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 104 from Patent WO02068623.
ACCESSION AX534866
VERSION AX534866.1 GI:25261452
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE
1
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A.,
Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W.,
Heinrich,O., Maier,D., Spreafico,F., Folkers,G., Hopper,S.,
Kemmer,W., Tan,P., Stiebler,J. and Albani,R.
TITLE Novel genes encoding novel proteolytic enzymes
JOURNAL Patent: WO 02068623-A 104 06-SEP-2002;
DSM N.V. (NL)
FEATURES
source
1..1665
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 374 a 455 c 424 g 412 t
ORIGIN
Query Match 7.4%; Score 118.6; DB 6; Length 1665;
Best Local Similarity 47.2%; Pred. No. 1.7e-19;
Matches 673; Conservative 0; Mismatches 649; Indels 105; Gaps 6;
Qy 121 GCGCTGTCGAGACTACCGGGTGTCAAACTCTCTGATATGTCGACACCTCTCC 180
Db 130 GGAATTTCCGAACTACCGGGTGTCCGATCTTAITCGGGTATGTACACCTTCCCGC 189
Qy 181 GAGTCCCATACCTCTCTGCTCTTCGAAGCAGACATAACCCAGAACTGACCTATC 240
Db 190 GCCTCACACGCTCTTTTGGTTTTTTCGAAGCCGCAAGATCCGACCAATGCGCTCTG 249
Qy 241 ACATTGGTGTGAATGGTGGCCCTGGAAGCGATTTTGTATCGGTCTTTCGAGAGTTG 300
Db 250 GCGATCTGCTCAATGGCGGTCCGGTGTGCTGCTCATGGGCTCTTGAAGAATA 309
Qy 301 GCGCTTGCATGCAATTCGACTTTTATGAC---TACATCAACCCCTCACTCGTGAAC 357
Db 310 GGTCTTTTCCATTCGATCAGATCCAGACACAGCTTCAATCTTGGAGTTGGAAC 369
Qy 358 GAGGTCTCCAAATTTACTATTCTGTCGCCAGGCAATTTGGGAGTGGCTTTTCATATAGT 417
Db 370 AATGAAGTCAATCTCTATCTTCTTGACCGACCACTCAAGTGGCTTCTCATAGATGC 429
Qy 418 ACGTGTGATGGTTCATTAACCTCTAATCGGGTGTGCGAAATTCGAGCTTTCAGGA 477
Db 430 CCAACAAATGSCACTTTGACAGCTA----- 454
Qy 478 GTTCAGGCGCGGTACCCCAACCATTTGATGCCACTCTGATCACTACTCAATCTTCCGCA 537
Db 455 -----ATGGACTGCTATTCGCGCT 474
Qy 538 GAGCGCGTGTGGAGATCTCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGG 597
Db 475 CACGCTCTATGCAATTTCCGCAACCTGTTTTCGAGTTCCCACTACAGCAAC 534
Qy 598 GTGAGTCTAGGACTTCACTATGAGGAGAGCTATGAGGCGCACTATGCTCTGCA 657
Db 535 GATGATCTG-----TCAGTCTCTGGCTGAAAGTTACGGAGGCAATTTGCTCAGGC 588
Qy 658 TCTTCAATCATTTTACGAGCAATGAGAGAAATGCAACGCTAGTGTGTTAATGCTGT 717
Db 589 ATCTTTCGCTCTTCCACAGCAGATGACAAATTCGAGAGGAGCTGCAAGACGCT 648

QY 718 -----CAGCTTAATTTAACTCTCTGGGAATTAATTAACGCGCATCATCGACGAGCGATC 771
Db 649 GCACAGTATTTGCACTCTCGACACGCTTGGCATTTGACGCGCTTCAATGATATGATGATC 708
QY 772 CAGCGCCCTTACACCTGATTCGCTGTGAACATACCTACGCTATCAAGCGTGTCAAC 831
Db 709 CAAGAAGAGGCTTACATCTACTTGCCCATACATA--ACGTAAGCGTCGCCCTTCTTCAT 766
QY 832 GAGACCGCTTACAACTACATGAAGTTTCCCAACCAATGCCAAAGTGTGCGCAAGTTG 891
Db 767 TCACTCGCAGGCTTTCGCGATCAGCGCTTCGCGCTCGAAGCGGCTTTGAAGAAGCGG 826
QY 892 ATTTCCACCTGCAACACAGAAACCGCAGCGCATTAAGTCTACTACGCGCTCTGCGCGAA 951
Db 827 ATTCCGGCTTGC-----TCACTCAGGGAAGATATCTCTGAATTTGGGA 873
QY 952 GCCACCAACATGTCAGGACATATGTTGAGGGGCATACACTACGCGCTTTCGCTGCGGT 1011
Db 874 GGCCTTGCACTAGAAATGGGAGATGGGCCCATACCTACTACCAACCTTCAATCGCGGG 933
QY 1012 GTGTATGATATCGGCATCATATGATGACCGGACTCGCCAACTTATACAAATTT 1071
Db 934 TGGTACGACATCGCCATCTAAGAACGCCATTCCTGCAAGACATGCTCGGAT 993
QY 1072 CTGCAAGGACTCTGTCTATGGAGCTATGGCGTCAACATCAACTACACCGAGTCCAA 1131
Db 994 TTGACGAGGAGTCCGCTTGGCGCTTGGGGTACCACTATTCACATCTGCTCTCG 1053
QY 1132 AATGACGCTTACAGCTTTCCAGCAAAACAGGCGACTTTGCTGGGCCAACTTCAATCGAA 1191
Db 1054 AGTGGCGTGTCTACAGATTTATATAAAACCTTTGATTCCTCCACGCGGCTTCCIGAT 1113
QY 1192 GACCTCGAGGAGATCTGTCTCTCCCGTGGGTGCTCCCTCTGCTGGAATCTACTCCCAAGCC 1311
Db 1114 GCATTTGGTACTCTCTGACAGTGGTGTAAAGTACACATGATACGGAGATCGTGAT 1173
QY 1252 TACATCTGCAACTGGTTCGCGCTCAGCGCGTTTCCCTGCTGGAATCTACTCCCAAGCC 1311
Db 1174 TAGCGCTGCAATTTGGTTCGCGGCGAAAGCCAGCTTTCAGTTCGATTCCTGATC 1233
QY 1312 GCCAGTTCGAGCGGAGGTACACGCCCTGAAAGTCAACGCGCTGAGTATGGGAA 1371
Db 1234 ACCGAATTTGCCACACGCGGATATCCCACTCTTACGCCCGACGGGATCAGCGCATG 1293
QY 1372 ACTCGGAGTATGTAATTTCTCTTCACTCGGCTCTATGAGCGAGCCATGAGTCCCA 1431
Db 1294 ACCCGCAGCTGGCAACTACAGCTTCACTCGCGCTTCCACGCGGATGAGGTCCTCC 1353
QY 1432 TACTACAGCCCATCGCTCCCTGCAATTTGTTTAAOCCGACTATCTTCGGTTGGGATC 1491
Db 1354 TCTACAGCGCTGTCGCGGATGATGATCTTCATCGGCGGACATTCACAAAGATAIC 1413
QY 1492 GCAGAGGCGCAGAAAGATCTGGCCCGACCTACAGCAAGATGGAAC 1538
Db 1414 CCTACTGGCTCTTGGCTGTGTGATGAGAAATCCACGTCGTTGGAC 1460

RESULT 7
AX534824
LOCUS AX534824 1581 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 62 from Patent WO02068623.
ACCESSION AX534824
VERSION AX534824.1 GI:25261366
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE 1
AUTHORS Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustadt Von,W., Heinrich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S.,

Kemmer,W., Tan,P., Stiebner,J. and Albang,R.
Novel genes encoding novel proteolytic enzymes
Patent: WO 02068623-A 62 06-SEP-2002;
DSM N.V. (NL)
Location/Qualifiers
1..1581
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
BASE COUNT 343 a 473 c 423 g 342 t
ORIGIN
Query Match 7.3%; Score 116.8; DB 6; Length 1581;
Best Local Similarity 52.5%; Pred.No.5.le-19;
Matches 334; Conservative 0; Mismatches 287; Indels 15; Gaps 3;
QY 914 ACCGACCGCATTTAGCTAGCTACGCGCTTCGCGGAAGCCACCAACATGTCGAGGACA 973
Db 812 ACTGGCGCGCCGAGGCATCGACGAGATCTGACGACTGCGCGACGATTTTTCGCGCAACG 871
QY 974 ATGTTGAGGGCCATACACTACGCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1033
Db 872 AGGTGAAAAGCTCTACGACATTTACTTCGGTTCGGGATGAGTACTGCTTTCGTAACATCA 931
QY 1034 ATGAIGACCGGACTCCGCCAAGTTTATACACAAATTTCTGCGAAAGAGCTCTGCTATGG 1093
Db 932 CTCGGGACCGCTTCCCTTACGAGTTCACGTTGACTACCTGAACAAAGCGTCCGTCGAG 991
QY 1094 ACGTATCGGGCTCAACATCACTACACCGGCTCAATGAGTCTACTACGCTTTC 1153
Db 992 CCGCCATCGGCGATACATCAATTTACAGGAGAGACAAACGCTGTTGGACTCGCCTTT 1051
QY 1154 AGCAACAGCGGACTTTGCTGCGCC---AATTTATGGAAGAGCTCGAGAGATCTGCTG 1210
Db 1052 CGTCCAGCGGTGACGAGCGGCGACTCATGAACACCATCCAGGATGTGGCAAGCTGCTCA 1111
QY 1211 CTCCTCCGCGGTGCTCTCCCTCATCTATGCGGAGCGGATATCATCTGCAACTGCTG 1270
Db 1112 AACAGGTGTACCGGTGTATGATACCGCGGAGTCCGACTAATACTGCAACTGCTGCTG 1171
QY 1271 CGGCTAGGCGGCTTTCCTCTGCTGCAACTACTCCCAAGCGCGCTTCCGAAGCGGAG 1330
Db 1172 GTGGGAGCGGCTG-----CGTTGAGGTCAAGCGCGCACTTCACTAGTGGG 1222
QY 1331 GGTACAGCGCC---CTGAAAGTCAACGCGCTGAGTATGCGGAACTCGCGAGTATGTA 1387
Db 1223 GTTACACCAACTTGTCACTCGGATGAGTACACACGCGGAGTTCGCGAGCGGCGG 1282
QY 1388 ATTTCTCTTCACTCGGCTATGAGCGAGCGGCTGAAGTCCCATCTACAGCCCATCG 1447
Db 1283 AATTGCTTTGTCGAGTGTATGAGAGTGGACATGAGTTCCTTCTATCAACCTTTC 1342
QY 1448 CTTCCCTGCAATTTTAAACCGGACTATCTTCGTTGGGATATCGCAGAGCGGCAAGA 1507
Db 1343 TTGCGCTGGAGATTTTGAAGCGCTCATTTGGGCGAGGATGTGCGACGGAAGATTC 1402
QY 1508 AGATCTGGCGCCAGCTACAGCAAGCAAGTGAAGCGCTA 1543
Db 1403 CCATCTGCTGAGTTTACAGACGCTGGGCGACGCCA 1438

RESULT 8
AX534767
LOCUS AX534767 2940 bp DNA linear PAT 22-NOV-2002
DEFINITION Sequence 5 from Patent WO02068623.
ACCESSION AX534767
VERSION AX534767.1 GI:25261241
KEYWORDS Aspergillus niger
SOURCE Aspergillus niger
ORGANISM Aspergillus niger
REFERENCE 1
AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

AUTHORS	Edens,L., dijk Van,A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gustedt Von,W., Heirich,O., Maier,D., Spreafico,F., Folkers,U., Hopper,S., Kemmer,W., Tan,P., Stiebler,J. and Albang,R.
TITLE	Novel genes encoding novel proteolytic enzymes
JOURNAL	Patent: WO 0206823-A 5 06-SEP-2002; DSM N.V. (NL)
FEATURES	Location/Qualifiers
source	1..2940
BASE COUNT	691 a 818 c 732 g 699 t
ORIGIN	
Query Match	7.3%; Score 116.8; DB 6; Length 2940;
Best Local Similarity	52.5%; Pred. No. 5.4e-19;
Matches	334; Conservative 0; Mismatches 287; Indels 15; Gaps 3;
QY	914 ACCGCACCGCATTAGCTGACTAGCGCCCTTCGCCCAAGCCACCACATGTCAGGGACA 973
Db	1492 ACTGCCGCCCGGAGCATCGAGAGCTGCGAGCACTGCGAGATTTTGGCGCAAG 1551
QY	974 ATGTTGAGGGGGCATACTACGCCCTTTCGTGGTGTGTATGATATTCGCATCCAT 1033
Db	1552 AGTTCGAAACGCTCAGGACATTTACTCCGGTCGGATGAGTAAGTTCGTGAACTCA 1611
QY	1034 ATGATGACCGGACTCCGCCAAGTTATACAAATAATTTCTGGCAAGGACTCTGTGATGG 1093
Db	1612 CTCGGACCCGTTCCCTTACGAGTTCTACGTTTACTACCTGAACAAGGCTCGCTGACG 1671
QY	1094 ACCTATCGCGGTCAACATCAACTACACGAGTCCCAATATGACGTCCTACTACGCTTTC 1153
Db	1672 CCGCCATCGGCGCATCATCATTTACGAGAGACACACGCTGTGTGGACTCGCCTTTT 1731
QY	1154 AGCAACAGCGGACTTTGTCTGGCCC---AACTTCATCGAAGACCTCGAGGAGATCCTTG 1210
Db	1732 CGTCCACCGGTGACGAGGGCGACTCATGAACACCATCCAGGATGTGGGCAAGCTGCTCA 1791
QY	1211 CTCTCCCGTGGCTGTCTCCTCATATGAGGCGAGCGCGATATCATCTGCAACTGGTTCG 1270
Db	1792 AAGAGGGTGTACGGTGGTTCATGACCGCGGGATGCCGACTATACTGCAACTGGCTGG 1851
QY	1271 GCGTCAAGCGGTTTCCCTCGCTCGGAACATCTCCCAAGCGGCCAGTTCGGAAGCGCAG 1330
Db	1952 GTGGGGAAGCGGTGT-----CGTTCAGGTTCAGGCCGCCCACTTCAGTAGTCGG 1902
QY	1331 GGTACAGGCC---CTGAAAGTCAACGGCGTGCAGTATGGGGAACATTCGGAGTATGGTA 1387
Db	1903 GTTACCAACAATTGTCACTCGSATGGATGGATGACACACGCGCAGGTGCGCGAGCGGGCG 1962
QY	1388 ATWTCTCCTTCACTCCGGTCTATGAGCGAGCCCATGAAGTCCCATACTACTACGCCCATCG 1447
Db	1963 AATTGCTCTTGTCCGAGTGTATGAGAGTGGCATGAGGTTCGCTTCTATCAAGCCCTTGC 2022
QY	1448 CCTCCCTGCAATTTGTTTAACCGGACTATCTTCGGTTGGGATATCGCAGAGGCCACAGA 1507
Db	2023 TTGGCTGGAGATGTTTGGCGGCTCATTTGGCGCAGGATGTGGCACGGGAAGATTC 2082
QY	1508 AGATCTGGCCCCAGCTACAAAGACGAATGGAACGGCTA 1543
Db	2083 CCATCTCGTGGAGTTTATACAGCGGTGGCGACGGCCA 2118

RESULT_9	AX534809	AX534809	Sequence 47 from Patent WO02068623.	3080 bp	DNA	linear	PAT 22-NOV-2002
LOCUS	AX534809						
DEFINITION	AX534809						
ACCESSION	AX534809						
VERSION	AX534809.1		GI:25261332				
KEYWORDS							
SOURCE	Aspergillus niger						
ORGANISM	Aspergillus niger						

[illegible]

RESULT 10				
AX534815				
LOCUS	AX534815	3221 bp	DNA	linear
DEFINITION	Sequence 53 from Patent WO02068623.			
ACCESSION	AX534815			
VERSION	AX534815.1			
KEYWORDS	. GI:25261346			
SOURCE	Aspergillus niger			
ORGANISM	Aspergillus niger			
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;			

Aspergillus niger
Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

1. *Ascomycetes, trichomycetes, mitosporic trichomycetes*; *Aspergillus*

REFERENCE

AUTHORS

Edens, L., dijk Van, A. A., Krubasik, P., Alberman, K., Stock, A., Kimpel, E., Klubauber, S., Wagner, C., Fritz, A., Gustedt Von, W., Heinrich, O., Maier, C., Spreier, C. F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiehler, J. and Albani, K.

TITLE

Novel genes encoding novel proteolytic enzymes

JOURNAL

Patent: WO 02068623-A 53 06-SEP-2002;
DSM N.V. (NL)

```

FEATURES
  source
    Location/Qualifiers
      1..3221
        /organism="Aspergillus niger"
        /mol_type="genomic DNA"
        /db_xref="taxon:5061"
  BASE COUNT
    739 a 739 c 869 g 840 t

```

Query Match	6.7%	Score 107.8;	DB 6;	Length 3221;
Best Local Similarity	48.5%	Pred. No. 1.2e-16;		
Matches 449;	Conservative	0;	Mismatches 437;	Indels 39;
				Gaps 4;

[illegible]

QY	674	ACGAGCAGAAATGAGAGAAATGCCAACGGTAGTGTTAAAT--GGTGTTCAGCTTAATTCA	730
Db	1582	AGGAGATGAATTCAGAGAGATTTCAGAGTGGGGAGGTAAACCCGGGAAGAAGATCCATTTGG	1641

QY 731 ACTCTGTGGGAATTATTAACGCAICATCGACGAGCGATCCAGGCCCTTACTACCCTG 790
| | | | |
Db 1642 ATACGCTGGGCATTATCAATGGTGTTGGATTACTCGTCGAGGTCCCTTCGTTCCCTG 1701

QY	791	AATTCGCTGTGAACAATACCTACGGTATCAAGCTGTGCACGAGACCGTCTACAACTACA	850
db	1702	ACGACGGCTAATACATTCGTATGGTGATCCGCCCATCTTTCCGCGCCTGCTG	1761

851 TGAAGTTGCCAACCAATGCCAAATGGTGGCAGGATTTGATTTCCACCTG-----902

DB	1/62	903	946
QY	CTATGGATAGTTGGAGCAAGGCTGGCGGGTGCAGGGAATATGATCATCGAGTSTCGCGATG	-----CAACAGACAAACCGCACCGCATTAGCTGACTACGCGCTCTGCG	-----CAACAGACAAACCGCACCGCATTAGCTGACTACGCGCTCTGCG

DB	1822	947	1881
CTGGGAGCTCGGAGATCCCTCATGTATGGCGACAATGAGACGGTAAATACCACTGCG	CCGAAGCCACCAACATGTGCAGGGGACAAAGTTGAGGGGCCATCTACTACGCCCTTGTCTGTC		

Db	1882	AGGAGGCGT	CGGACTACTGTT	CGGAGATCAAGAGCCTG	TATACGAATACCTCCGGC	1941
QY	1007	GTGGTG	TGATGATTCGGATCCAT	TATGATACCGGACTCCG	CAAGTTATTACACA	1066

D _b	1942	GAGGATAC	TACGACAT	TAGGCGATT	TTCACGCCG	GATCGAGCT	CTCGTGCCT	TACTTCGTG	2001
Q _y	1067	AATTCTTGG	CAAAAGGAC	TCTCTCATG	CAAGCTAT	CGCGCTCA	CACTCAAC	TACACCAGT	1126

Db	2002	GGTCTTGATCGCCCATGGGTGCAAGGCACCTTGGGTCCCGGTGAACATATACCATGT	2061
Oy	1127	CCAAATAATGACGTCTACTACGGTTTCGACGAACAGCGCAGT-----TGTCTGTCG	1177

Db	2062	CGTCAGGCGCASTGGGGAACGATTTCGCTCGACGGCGATTATCCGCGAAATGATCCCC	2121
QY	1178	CCAACATCAATCGGAAGACCTCGAGGAGATCCTTGCTCCCGCTGGGTCTCCCTCATCT	1237

[illegible]

2182 ATGGGACCGGACATGCTTGCCGTGCCGGCGGGGAAGATGCAGCCCTGCTGTGG 2241

1296	ACACACCCAAAGCCGCCCTTCGGAAGCGGAGGTCACGCCCTCGAAAGTCACGGCG	1355
1297		1356
1298		1357
1299		1358
1300		1359
1301		1360
1302		1361
1303		1362
1304		1363
1305		1364
1306		1365
1307		1366
1308		1367
1309		1368
1310		1369
1311		1370
1312		1371
1313		1372
1314		1373
1315		1374
1316		1375
1317		1376
1318		1377
1319		1378
1320		1379
1321		1380
1322		1381
1323		1382
1324		1383
1325		1384
1326		1385
1327		1386
1328		1387
1329		1388
1330		1389
1331		1390
1332		1391
1333		1392
1334		1393
1335		1394
1336		1395
1337		1396
1338		1397
1339		1398
1340		1399
1341		1400
1342		1401
1343		1402
1344		1403
1345		1404
1346		1405
1347		1406
1348		1407
1349		1408
1350		1409
1351		1410
1352		1411
1353		1412
1354		1413
1355		1414
1356		1415
1357		1416
1358		1417
1359		1418
1360		1419
1361		1420
1362		1421
1363		1422
1364		1423
1365		1424
1366		1425
1367		1426
1368		1427
1369		1428
1370		1429
1371		1430
1372		1431
1373		1432
1374		1433
1375		1434
1376		1435
1377		1436
1378		1437
1379		1438
1380		1439
1381		1440
1382		1441
1383		1442
1384		1443
1385		1444
1386		1445
1387		1446
1388		1447
1389		1448
1390		1449
1391		1450
1392		1451
1393		1452
1394		1453
1395		1454
1396		1455
1397		1456
1398		1457
1399		1458
1400		1459
1401		1460
1402		1461
1403		1462
1404		1463
1405		1464
1406		1465
1407		1466
1408		1467
1409		1468
1410		1469
1411		1470

Qy	1358	TCGAGIATGGGG---AAACTCGCAGTATGTAATTTCTCCTTCACTCGGGTCTATGAGG	1414
Db	2302	CTACGTTGGGGTCTAGIRAGGCAGTATGGACATCTCTGTTCCACGGTGTCTTTCAGG	2361
Qy	1415	CAGCCCATAGTCCCATCTACACGCCCATCGCCTCCCTGCAATTTGTTAAACCGGACTA	1474
Db	2362	CGGCCCATAGGTGCCATTTTATCAGCCCGAAAGCGCGTATGAGATTTTAAATCGCGCTC	2421
Qy	1475	TCCTCGTTGGGATATCGCAGGG	1499
Db	2422	AGTTAATTTGGGATATTGCGACGGG	2446

RESULT 11		
A75535		
LOCUS	A75535	2503 bp
DEFINITION	Sequence 2 from Patent	DNA
ACCESSION	A75535	linear
VERSION	A75535.1	PAT 15-OCT-1999
	GI:6065609	

KEYWORDS
SOURCE
ORGANISM
unidentified
unidentified
unclassified.

REFERENCE
1 (bases 1 to 2503)
AUTHORS
Fleer, R., and Fournier, A.
TITLE
MODIFIED KLUYVEROMYCES YEASTS. THEIR PREPARATION AND USE
JOURNAL
BIOLOGICAL WORTS, 94, 00579-83, 05787, 1990

SOURCE JOURNAL OF DOCUMENTATION, vol. 49, no. 2, JAN 1994, pp. 107-116, 18 refs.

FEATURES RHONE POULENC PAPER SA (FR); FLEER REINHARD (FR)
Location/Qualifiers
1..2503

```

/organism="unidentified"
/mol_type="genomic DNA"
/strain="Kluiveromyces lacticus"
/db_xref="taxon:32644"

```

```
gene
387. .1862
      /gene="K1_PRC1"
CDS
387. .1862
      /gene="K1_PRC1"
```

```

/Codon_start=1
/product="Gene de la protease C de K.lactis"
/protein_id="CAB5852.1"
/feature_id="CAB5852.1"

```

```

/translation: "MVSIRKELLSYGLWSVTALSLNVAVDSLSFNSFCGNNTTDEH
TANYNTQFVSFNSSNDIDGSLRIKPLDPSGVDVTKQWSGYLDYQDSRHF.YWFFES
RNDYNTQFVILNMGDGGCSVDFELPGLSGADLPIYNYFWSNAGNSVFLDQ
/translation: "GL:0083610"

```

PVGVEFSIGDGSKSTIDDAARDVIFLDFLFEFEHLRHHFHSVESXAGHYLPKII
 REIIVAAHEDSKSVSLVJLNGNDFDPLAQYQYEPMACGEGFPAVLVEPCLDMNR
 NNPLCLSLVRCYKSHSVFCLADRYCEQQTGVYIEKSRNPDPISKEAEADSGA
 CYQEELIISDLYNQEVQRALGTVDVSSFGQCSDDVGFATGTPSPHQYVAELLD
 PVGVEFSIGDGSKSTIDDAARDVIFLDFLFEFEHLRHHFHSVESXAGHYLPKII

BASE COUNT
ORIGIN

	703 a	475 c	517 g	808 t
YGPILYLRIYDAGHNVPHDPQENSLQMVNSIQNLAKRSRI"				
ODINVLIVAGDKDYICNWLGNLAWTEKLEWRYNFYKKQVLRTPKSEEDETIGEIKR				

Query Match	5.4%	Score 86.4;	DB 6;	Length 2503;
Best Local Similarity	57.4%	Pred. No. 4e-11;		
Matches 156:	Conservative	0;	Mismatches 116;	Indels 0;

[illegible]

DG	624	G T G A A C A A A T G G C G G G A I A T T T A G A T T A C C A G G A C T C A A A A C A C T T C T T T T A T T G G T T T	683
Qy	205	T T C G A A C C C A G A C A T A A C C C G A A A C T G C A C T A T C A C A I T T G T G T T G A A T G G T G G C C C C T	264

Db	684	265
QY	TTTGAGTCTAGAAAATGACCCAGAGAAATGACCGAGTACTATGGTTAAACGGTGGTCT	GGAAGGCAATCTTTGATCGGGTCTCTCCGAGAGTTGGGCCCTTGCCATGTCAATTGCACT

Dδ	744	GSCGTGCCCTTTTCGTGCGGCTTTTCTTGAATIGSGACCTTCTCTATAGGACTGAT	803
QY	325	TTTGATCACTACATCACCTCCTCCTCGTGGACGAGTCTCCAATTACTTAITCCTGTCC	384

Db
804 TTGAACCCCTTTATACCCCTACTCTTGGAAATCCCAACGCTTCGTGATATCCCTAGAT 863


```

QY 385 CAGCCATTGGAGTCGGCTTTTCATATAGTGA 416
||||| ||||| ||||| ||||| |||||
Db 864 CAGCCATTGGAGTCGGCTTTTCATATAGTGA 895

RESULT 12
I70282
LOCUS I70282 2503 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 3 from patent US 5679544.
ACCESSION I70282
VERSION I70282.1 GI:3006417
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2503)
AUTHORS Fleer, R., Fournier, A. and Yeh, P.
TITLE Modified Kluyveromyces yeasts, their preparation and use
JOURNAL Patent: US 5679544-A 3 21-OCT-1997;
FEATURES
Location/Qualifiers
source 1..2503
/organism="unknown"
BASE COUNT 703 a 475 c 517 g 808 t
ORIGIN

Query Match 5.4%; Score 86.4; DB 6; Length 2503;
Best Local Similarity 57.4%; Pred. No. 4e-11;
Matches 156; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 145 GTCAAATCTTACCTGATATGCGACACCTCTCCGAGTCCCATACCTTCTTCTGTTTC 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 624 GTGAACAATGGTCGGGATATTAGATTACAGGACTCAAAACACTCTTTTATGTTT 683

QY 205 TTGGAAGCCAGACATACCCAGAACTGCACCTATACATGTGTTGATGTTGGCCCT 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 684 TTGAGCTAGAAATACCCAGAGAATGACCCAGTATAGTTGTTAAACGGTGGTCT 743

QY 265 GGAAGGATCTTTGATCGTCTCTCGAAGAGTTGGCCCTGCCATGTCATTCGACT 324
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 744 GGCTGTTCTCTTCTGTCGGTCTTTCTTTGAAATGGGACCTTCTTCTATAGGACGAT 803

QY 325 TTGATGACATACATCAACCTCTCTGTTGAAGAGGCTCTCCAAATTTTACTTCTCTGTC 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 804 TTGAACCCATTTATTAACCCCTCTCTGGAATTCACACCTCTCTGATATTCCTAGAT 863

QY 385 CAGCCATTGGAGTCGGCTTTTCATATAGTGA 416
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 864 CAGCCATTGGAGTCGGCTTTTCATATAGTGA 895

RESULT 13
I33983
LOCUS I33983 2002 bp DNA linear PAT 06-FEB-1997
DEFINITION Sequence 3 from patent US 5594119.
ACCESSION I33983
VERSION I33983.1 GI:1824774
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2002)
AUTHORS Yaver, D.S. and Thompson, S.A.
TITLE Gene encoding carboxypeptidase of aspergillus niger
JOURNAL Patent: US 5594119-A 3 14-JAN-1997;
FEATURES
Location/Qualifiers
source 1..2002
/organism="unknown"
BASE COUNT 416 a 591 c 505 g 490 t
ORIGIN

Query Match 5.2%; Score 84; DB 6; Length 2002;
Best Local Similarity 55.5%; Pred. No. 1.6e-10;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACCTCTCCGAGTCCCATACCTTCTTCTGGTTCCTCGAAGCCACACATAACCCAGAA 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 617 GACAAAGAGAACGACAAAGCATCTGTCTTCTGAGTCTCGAGTCTCGCAATGACCCGAG 676

QY 229 ACTGCACCTATACATTTGTTGATGTTGGTGGCCCTCGAAGCGATTTCTTGTATCGGTCTC 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 ATGACCTGTGTTCTGTGCTGACGGTGGCCCTGAGTCTTCCCTCACCGTCTT 736

QY 289 TTCGAAGAGTTGGCCCTTGCATGTCATTTGATGACITTTGATGACITACATCAACCCCTAC 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 TTCATGGAGTCGCGCCCTAGCAGCATCAACAAGAATCCAGCGGCTTACAGCACTAC 796

QY 349 TCGTGGAAAGAGGTCTCCAAATTTACTATTCTCTGCCAGCATTTGGGAGTCGGTCTTCA 408
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 GCTTGGAACTCCAAACGGTCCGTGATCTTCTTGACCAGCTGTCAACGTCGGTACTCT 856

QY 409 TATAGTGATACGGTTGATGGTCCATTAACCTGTACTGGGGTGGTCGCA 458
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 857 TACAGCAACCTCTGCTGTACGAGACACCGTTGCTGCTGGCAAGGACGCTCTA 906

RESULT 15
I77239
LOCUS I77239 2002 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 3 from patent US 5693510.
ACCESSION I77239

```

```

Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACCTCTCCGAGTCCCATACCTTCTTCTGGTTCCTCGAAGCCACACATAACCCAGAA 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 617 GACAAAGAGAACGACAAAGCATCTGTCTTCTGAGTCTCGAGTCTCGCAATGACCCGAG 676

QY 229 ACTGCACCTATACATTTGTTGATGTTGGTGGCCCTCGAAGCGATTTCTTGTATCGGTCTC 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 ATGACCTGTGTTCTGTGCTGACGGTGGCCCTGAGTCTTCCCTCACCGTCTT 736

QY 289 TTCGAAGAGTTGGCCCTTGCATGTCATTTGATGACITTTGATGACITACATCAACCCCTAC 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 TTCATGGAGTCGCGCCCTAGCAGCATCAACAAGAATCCAGCGGCTTACAGCACTAC 796

QY 349 TCGTGGAAAGAGGTCTCCAAATTTACTATTCTCTGCCAGCATTTGGGAGTCGGTCTTCA 408
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 GCTTGGAACTCCAAACGGTCCGTGATCTTCTTGACCAGCTGTCAACGTCGGTACTCT 856

QY 409 TATAGTGATACGGTTGATGGTCCATTAACCTGTACTGGGGTGGTCGCA 458
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 857 TACAGCAACCTCTGCTGTACGAGACACCGTTGCTGCTGGCAAGGACGCTCTA 906

RESULT 14
I74375
LOCUS I74375 2002 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 3 from patent US 5688663.
ACCESSION I74375
VERSION I74375.1 GI:3010516
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2002)
AUTHORS Yaver, D.S. and Thompson, S.A.
TITLE Gene encoding carboxypeptidase of Aspergillus niger
JOURNAL Patent: US 5688663-A 3 18-NOV-1997;
FEATURES
Location/Qualifiers
source 1..2002
/organism="unknown"
BASE COUNT 416 a 591 c 505 g 490 t
ORIGIN

Query Match 5.2%; Score 84; DB 6; Length 2002;
Best Local Similarity 55.5%; Pred. No. 1.6e-10;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACCTCTCCGAGTCCCATACCTTCTTCTGGTTCCTCGAAGCCACACATAACCCAGAA 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 617 GACAAAGAGAACGACAAAGCATCTGTCTTCTGAGTCTCGAGTCTCGCAATGACCCGAG 676

QY 229 ACTGCACCTATACATTTGTTGATGTTGGTGGCCCTCGAAGCGATTTCTTGTATCGGTCTC 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 ATGACCTGTGTTCTGTGCTGACGGTGGCCCTGAGTCTTCCCTCACCGTCTT 736

QY 289 TTCGAAGAGTTGGCCCTTGCATGTCATTTGATGACITTTGATGACITACATCAACCCCTAC 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 TTCATGGAGTCGCGCCCTAGCAGCATCAACAAGAATCCAGCGGCTTACAGCACTAC 796

QY 349 TCGTGGAAAGAGGTCTCCAAATTTACTATTCTCTGCCAGCATTTGGGAGTCGGTCTTCA 408
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 GCTTGGAACTCCAAACGGTCCGTGATCTTCTTGACCAGCTGTCAACGTCGGTACTCT 856

QY 409 TATAGTGATACGGTTGATGGTCCATTAACCTGTACTGGGGTGGTCGCA 458
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 857 TACAGCAACCTCTGCTGTACGAGACACCGTTGCTGCTGGCAAGGACGCTCTA 906

RESULT 15
I77239
LOCUS I77239 2002 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 3 from patent US 5693510.
ACCESSION I77239

```


Search completed: September 16, 2003, 18:04:51
Job time : 7747 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 14:42:53 ; Search time 148 Seconds
(without alignments)
4795.568 Million cell updates/sec

Title: US-09-712-338-1_COPY_55_1662

Perfect score: 1608

Sequence: 1 ctccaggaagtacacggc.....gcattgtccagtgttggtatg 1608

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 559978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
 - 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
 - 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
 - 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
 - 5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1608	100.0	1662	3	US-08-943-714-1
2	86.4	5.4	2503	1	US-09-640-305-3
3	86.4	5.4	2503	1	US-08-360-673-3
4	84	5.2	2002	1	US-08-309-341-3
5	84	5.2	2002	1	US-08-608-267-3
6	84	5.2	2002	1	US-08-608-452-3
7	84	5.2	2002	1	US-08-608-224-3
8	84	5.2	2002	2	US-08-967-149-3
9	76	4.7	2068	1	US-08-309-341-1
10	76	4.7	2068	1	US-08-608-267-1
11	76	4.7	2068	1	US-08-608-452-1
12	76	4.7	2068	1	US-08-608-224-1
13	76	4.7	2068	2	US-08-967-149-1
14	66.2	4.1	2632	2	US-08-899-324-32
15	66.2	4.1	2632	3	US-08-329-892B-32
16	56.4	3.5	1814	4	US-09-702-705-319
17	56.4	3.5	1814	4	US-09-736-457-319
18	47.6	3.0	390	3	US-09-197-649-7
19	47.6	3.0	1551	2	US-08-828-488-4
20	47.6	3.0	1551	4	US-09-299-689A-4
21	47.2	2.9	1670	2	US-08-828-488-2
22	47.2	2.9	1670	4	US-09-299-689A-2
23	39.8	2.5	357	4	US-09-280-116-131
24	38.6	2.4	447	4	US-09-252-991A-11967
25	38.6	2.4	630	4	US-09-252-991A-8154
26	38.6	2.4	774	4	US-09-252-991A-11893
27	38.6	2.4	777	4	US-09-252-991A-8222

28	38.6	2.4	1083	4	US-09-252-991A-11930
29	38.6	2.4	1239	4	US-09-252-991A-8276
30	38.6	2.4	1434	4	US-09-252-991A-8046
31	38.2	2.4	918	3	US-09-105-390-35
32	38.2	2.4	1008	3	US-09-105-390-51
33	38.2	2.4	1949	3	US-09-105-390-2
34	37.2	2.3	870	4	US-09-252-991A-10938
35	37.2	2.3	1785	4	US-09-252-991A-11178
36	37.2	2.3	2091	4	US-09-252-991A-11076
37	36.8	2.3	7012	4	US-09-221-017B-173
38	36.4	2.3	1269	4	US-09-252-991A-8698
39	36.4	2.3	1443	4	US-09-252-991A-3554
40	36.4	2.3	1830	4	US-09-252-991A-3563
41	36.4	2.3	1926	4	US-09-252-991A-3531
42	36.4	2.3	1944	4	US-09-252-991A-9081
43	36.4	2.3	3494	3	US-09-139-802-200
44	36.4	2.3	3494	4	US-09-659-786-200
45	36.4	2.3	7218	1	US-08-232-463-14

ALIGNMENTS

RESULT 1

US-08-943-714-1
; Sequence 1, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION:
; APPLICANT: Blinkovsky, Alexander
; APPLICANT: Berkla, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Dambmann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 61875780 No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,714
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4990.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1662 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..1662
; OTHER INFORMATION:

100%

US-08-943-714-1

Query Match 100.0%; Score 1608; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTCCAGGAGTACCGGGGTCGGTCTGGTAGAGACAGCTACCCCAAGAACCCACCGGG 60
DB 55 CTTCCAGGAGTACCGGGGTCGGTCTGGTAGAGACAGCTACCCCAAGAACCCACCGGG 114

QY 61 GTCAGAGCTCTTCAACCGCAACATGTCACCATCGGTACAGAGAACCCGGGGCAGAG 120
DB 115 GTCAAGACTCTTCAACCGCAACATGTCACCATCGGTACAGAGAACCCGGGGCAGAG 174

QY 121 GGGGTCTGGCAGACTACCCGGGTGTCAAATCTCTCTGATATGTCGACACTCTCCC 180
DB 175 GGGGTCTGGCAGACTACCCGGGTGTCAAATCTCTCTGATATGTCGACACTCTCCC 234

QY 181 GAGTCCCATACCTTCTTCTGTTCTTTCGAAGCCAGACATAACCCAGAAACTGCACTTATC 240
DB 235 GAGTCCCATACCTTCTTCTGTTCTTTCGAAGCCAGACATAACCCAGAAACTGCACTTATC 294

QY 241 ACATTTGTTGTAATGTTGGCCCTGGAGCGATTTTGTATCGTCTCTTCGAAGAGTTG 300
DB 295 ACATTTGTTGTAATGTTGGCCCTGGAGCGATTTTGTATCGTCTCTTCGAAGAGTTG 354

QY 301 GGGCCTTGCCATGTCATTCGACTTTTGTGATGACTACATCAACCTCTACTGTTGGAACGAG 360
DB 355 GGGCCTTGCCATGTCATTCGACTTTTGTGATGACTACATCAACCTCTACTGTTGGAACGAG 414

QY 361 GTCTCCCAATTTACTTCTTCTGTTCTTTCGAAGCCAGACATAACCCAGAAACTGCACTTATC 420
DB 415 GTCTCCCAATTTACTTCTTCTGTTCTTTCGAAGCCAGACATAACCCAGAAACTGCACTTATC 474

QY 421 GTTGATGGTCCATTAACCTCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCT 480
DB 475 GTTGATGGTCCATTAACCTCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCT 534

QY 481 CAGGGCCGGTACCCCAACCAATGATGCCACTCTGATGACTATGAGGGGCTTTCATATAGTATAG 540
DB 535 CAGGGCCGGTACCCCAACCAATGATGCCACTCTGATGACTATGAGGGGCTTTCATATAGTATAG 594

QY 541 GCGCCTTGGGAGATCCCTGCAAGAGATTCCTTAGTGGACTACCTAGCTGGACTCTAGGCTG 600
DB 595 GCGCCTTGGGAGATCCCTGCAAGAGATTCCTTAGTGGACTACCTAGCTGGACTCTAGGCTG 654

QY 601 CAGTCTAAGGACTTCAGTCTATGCGGAGAGCTATGAGGGGCTATGAGGGGCTATGAGGGGCT 660
DB 655 CAGTCTAAGGACTTCAGTCTATGCGGAGAGCTATGAGGGGCTATGAGGGGCTATGAGGGGCT 714

QY 661 TTCAATCATTTTACGAGCAGAAATGCAAGAAATGCAAGAAATGCAAGAAATGCAAGAAATGCAAG 720
DB 715 TTCAATCATTTTACGAGCAGAAATGCAAGAAATGCAAGAAATGCAAGAAATGCAAGAAATGCAAG 774

QY 721 CTTAATTTCAACTCTCTGGGAAATTAATTAACGGCATCATPCAGAGGGGATCCAGGCGCCT 780
DB 775 CTTAATTTCAACTCTCTGGGAAATTAATTAACGGCATCATPCAGAGGGGATCCAGGCGCCT 834

QY 781 TACTACCTGTAATTCGCTGTGAACATACCTTACGGTATCAGGGTGTCAACGAGACCGTTC 840
DB 835 TACTACCTGTAATTCGCTGTGAACATACCTTACGGTATCAGGGTGTCAACGAGACCGTTC 894

QY 841 TACAACATGTAAGTTTGGCAACCAATGCTGCAAGAAATGCTGCAAGAAATGCTGCAAGAAATGCT 900
DB 895 TACAACATGTAAGTTTGGCAACCAATGCTGCAAGAAATGCTGCAAGAAATGCTGCAAGAAATGCT 954

QY 901 TGCAAAACAGACAAACCGCAGCATTAAGTCTAGTACGCGCTCTGCGGCGGAGGCGGAGCCAC 960
DB 955 TGCAAAACAGACAAACCGCAGCATTAAGTCTAGTACGCGCTCTGCGGCGGAGGCGGAGCCAC 1014

QY 961 ATGTGCGGGGACAAATGTTGAGGGGCTACTAGCGCTTGTGTTGGTGGTGGTGGTGGTGGTGGT 1020
DB 1015 ATGTGCGGGGACAAATGTTGAGGGGCTACTAGCGCTTGTGTTGGTGGTGGTGGTGGTGGTGGT 1074

QY 1021 ATTGGCATCATATGATGACCCGACTCCGCAAGTTTATTACAAATTTCTGGCAAG 1080
DB 1075 ATTGGCATCATATGATGACCCGACTCCGCAAGTTTATTACAAATTTCTGGCAAG 1134

QY 1081 GACTCTCTGATGAGCGGTATCGGGTCAACATCAACTACACCCAGTCCATTAATGAGTTC 1140
DB 1135 GACTCTCTGATGAGCGGTATCGGGTCAACATCAACTACACCCAGTCCATTAATGAGTTC 1194

QY 1141 TACTACCTTTCAGCAAAACAGCGACTTTTGTCTGGCCCACTTCATCTGGAAGACTCGAG 1200
DB 1195 TACTACCTTTCAGCAAAACAGCGACTTTTGTCTGGCCCACTTCATCTGGAAGACTCGAG 1254

QY 1201 GAGATCTCTGCTCTCCCGCTGCTCTCCCTCATCTATGCGCGAGCGCGATTAATCTGTC 1260
DB 1255 GAGATCTCTGCTCTCCCGCTGCTCTCCCTCATCTATGCGCGAGCGCGATTAATCTGTC 1314

QY 1261 AACTGGTTCGGCGGTAGCGCGTTCCTCTCGCTCGCACTACTCCGAGCGCGCGCTTC 1320
DB 1315 AACTGGTTCGGCGGTAGCGCGTTCCTCTCGCTCGCACTACTCCGAGCGCGCGCTTC 1374

QY 1321 CGAAGCGCAGGGTACACGCCCTGAAAGTCAACGCGCTGAGTATGGGAAACTCGCGAG 1380
DB 1375 CGAAGCGCAGGGTACACGCCCTGAAAGTCAACGCGCTGAGTATGGGAAACTCGCGAG 1434

QY 1381 TATGTAATTTCTCTTCTACTCGCGTCTATGAGCGAGCGCTATGAGTCCCTTACTACAG 1440
DB 1435 TATGTAATTTCTCTTCTACTCGCGTCTATGAGCGAGCGCTATGAGTCCCTTACTACAG 1494

QY 1441 CCAATCGCTCCCTGCAATTTGTTAACCGGACTATCTTCGTTGGGATATCGCAGAGGCT 1500
DB 1495 CCAATCGCTCCCTGCAATTTGTTAACCGGACTATCTTCGTTGGGATATCGCAGAGGCT 1554

QY 1501 CAGAAGAGATCTGGCCAGCTACAGAGCAATGGAAGCGCTACAGCTACGATACACAG 1560
DB 1555 CAGAAGAGATCTGGCCAGCTACAGAGCAATGGAAGCGCTACAGCTACGATACACAG 1614

QY 1561 TCGTCCGTGCGGTGCGCTACGCTACGAGCATGTCAGTGTGGTATG 1608
DB 1615 TCGTCCGTGCGGTGCGCTACGCTACGAGCATGTCAGTGTGGTATG 1662

RESULT 2

US-09-640-305-3
; Sequence 3, Application US/09640305
; Patent No. RE37447
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/640,305
; FILING DATE: 16-Aug-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE: 06-FEB-1995
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993

```

RESULT 3
US-08-360-673-3
: Sequence 3, Application US/08360673
: Patent No. 5679544
: GENERAL INFORMATION:
: APPLICANT: Fleer, Reinhard
: APPLICANT: Fournier, Alain
: APPLICANT: Yeh, Patrice
: TITLE OF INVENTION: MODIFIED KLOYVEROMYCES YEASTS, THEIR
: TITLE OF INVENTION: PREPARATION AND USE
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rhone-Poulenc Rorer Inc.
: STREET: 500 Arcola Rd. 3043
: CITY: Collegeville
: STATE: PA
: COUNTRY: USA
: ZIP: 19002
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible

```

Query Match	5.4%;	Score 86.4;	DB 1;	Length 2503;
Best Local Similarity	57.4%;	Pred. No. 5.9e-16;		
Matches 156;	Conservative 0;	Mismatches 116;	Indels 0;	Gaps 0;
QY	145	GTCAAAATCTACTGTGATATGTGGACACCTCTCCCGAGTCCCATACCTTTCTTCGTTC	204	
Db	624	GTGAACAATATGGTCGGGATATTTAGATTACAGGACTCAAACACTTCTTTATTGGTIT	683	
QY	205	TTGGAAGCCGACATATACCCAGAAATCGACCTATCATCATGTGCTGAATGGTGGCCCT	264	
Db	684	TTTGAGTCTAGAATATACCCAGAAATGACCCAGTGAATCTATGGTTAAACGGTGGTCT	743	
QY	265	GGAAGCGATCTTTGATCGGTCTCTTCGAAGAGTGGGCCCTTCCCATGTCAATTCGACT	324	
Db	744	GGCTGTCTCTTTGTCGGTCTTTTCTTTGAATTTGGACCTTCTTCTATAGGAGCTGAT	803	
QY	325	TTTGATGACIATATCAACCTCTACTCTGGGAAGGAGGTCTCCAAATTTACTATCTCTGCC	384	
Db	804	TTGAACCCCATTTAAACCCCTACTCTTGAATTCCAACGCTTCGTGTGATATCTCTAGAT	863	
QY	385	CAGCCATGGGAGTCGGCTTTTTCATATAGTGA	416	
Db	864	CAGCCGTGGTGGTGTGGTCTCTCAIACGGTGA	895	

```

RESULT 4
US-08-309-341-3
; Sequence 3, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119o No. 5594119disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400

```

[illegible]

```

; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5693510o No. 5693510disk of No. 5693510th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,452
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; TELECOMMUNICATION INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
; US-08-608-452-3

```

```

Query Match          5.2%; Score 84; DB 1; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.8e-15;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACCTCTCCCGAGTCCCATACCTTCTCTGGTTCCTCGAAGCCAGACATAACCCAGAA 228
DB 617 GACAACGAGAACGACAGCATCTCTCTAGTTCCTCGAGTCTCGAATGACCCCGAG 676

QY 229 ACTGCACCTATCACATTTGTTGAATGGTGGCCCTGGAACGATCTTTGATCGGTCTC 288
DB 677 AATGACCTGTGTCTGTGCTGAGTGAACGGTGGCCCTGATCTCTTCCCTCACCGGCTT 736

QY 289 TTCGAAGAGTGGGCCCTTCGCATGTCAATTCGACTTTTGATGACTACATCAACCCCTAC 348
DB 737 TTCATGAGCTTCGCCCTTAGCAGCATCAACAAGAAGATCCAGCCGGTCTACAAAGACTAC 796

QY 349 TCGTGGACAGAGTCTCCAAATTTACTATTCCTGTCGCCAGCCATTTGGGAGTCGGCTTCA 408
DB 797 GCTTGAACATCCACGGCTCGGTGATCTCTCTGACCAAGCTGTCACAGTCGGTACTCT 856

QY 409 TATAGTATAGGTTGATGGTCCATTAACCCCTTAACCTGAGTGGGTCGTCGA 458
DB 857 TACAGCAACTCTGCTGTGTCAGGACACACCGCTTCTGCTGCGCAAGGAGCTCTA 906

```

RESULT 7

US-08-608-224-3

; Sequence 3, Application US/08608224

```

; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5705376o No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
; US-08-608-224-3

```

```

Query Match          5.2%; Score 84; DB 1; Length 2002;
Best Local Similarity 55.5%; Pred. No. 2.8e-15;
Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 169 GACACCTCTCCCGAGTCCCATACCTTCTCTGGTTCCTCGAAGCCAGACATAACCCAGAA 228
DB 617 GACAACGAGAACGACAGCATCTCTCTAGTTCCTCGAGTCTCGAATGACCCCGAG 676

QY 229 ACTGCACCTATCACATTTGTTGAATGGTGGCCCTGGAACGATCTTTGATCGGTCTC 288
DB 677 AATGACCTGTGTCTGTGCTGAGTGAACGGTGGCCCTGATCTCTTCCCTCACCGGCTT 736

QY 289 TTCGAAGAGTGGGCCCTTCGCATGTCAATTCGACTTTTGATGACTACATCAACCCCTAC 348
DB 737 TTCATGAGCTTCGCCCTTAGCAGCATCAACAAGAAGATCCAGCCGGTCTACAAAGACTAC 796

QY 349 TCGTGGACAGAGTCTCCAAATTTACTATTCCTGTCGCCAGCCATTTGGGAGTCGGCTTCA 408
DB 797 GCTTGAACATCCACGGCTCGGTGATCTCTCTGACCAAGCTGTCACAGTCGGTACTCT 856

QY 409 TATAGTATAGGTTGATGGTCCATTAACCCCTTAACCTGAGTGGGTCGTCGA 458
DB 857 TACAGCAACTCTGCTGTGTCAGGACACACCGCTTCTGCTGCGCAAGGAGCTCTA 906

```

```
RESULT 8
US-08-967-149-3
; Sequence 3, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59393050 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 28-FEB-1996
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 202 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
; US-08-967-149-3
; Query Match 5.2%; Score 84; DB 2; Length 2002;
; Best Local Similarity 55.5%; Pred. No. 2.8e-15;
; Matches 161; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 169 GACACTCTCCGAGTCCCATACCTCTCTGTTCTTGGAGCCAGACATAACCCAGAA 228
DB 617 GACACGAGAGACGAGCAAGCAATCTGTCTACTGGTCTTGGAGTCTGCAATGACCCCGAG 676
QY 229 ACTGCACCTATCATATGTTGGTGAATGGTGCCTGGAAGCAATCTTTGATCGCTC 288
DB 677 ATGACCCCTGTTGTTCTGGTGTACGCTGACGCTGCGCTGTGATGCTTCCCTCACCGGCTT 736
QY 289 TTGGAAGAGTTGGGCGCTTGGCAATTCGAATTCGACTTTTGTGATGACTACATCAACCTCAC 348
DB 737 TTATGGAGCTCGGCGCTAGCAGCATCAACAAGAAGATCCAGCGGCTCTACACGACTAC 796
QY 349 TCGTGGAGAGGAGTCTCCCAATTACTATTCTCTGTCCAGCCATGGGAGTCGGCTTTTCA 408

Db 797 GCTTGAACCTCCACACGCGTCCCGTGATCTTCTTGACCAAGCTGTCAACGTCGGTACTCT 856
QY 409 TATAGTGATACGGTTGATGGGTCCCAATTAACCCCTGAACCTGGGGTCTGCTGA 458
Db 857 TACAGCAACTCTGCTGTCAGGACACACGCGTGTGCTGCGGCAAGGACGCTCA 906

RESULT 9
US-08-309-341-1
; Sequence 1, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119o No. 5594119disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (571..633)
; US-08-309-341-1
; Query Match 4.7%; Score 76; DB 1; Length 2068;
; Best Local Similarity 55.8%; Pred. No. 7.5e-13;
; Matches 145; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
QY 199 TGGTCTTTCGAGCCAGACATAACCCAGAACTGCACCATATCACATTTGTTGATGTT 258
DB 633 TGGTCTTTCGAGTCTCGCAATGACCCCGAGAAATGATCCCGTTGTTCTGTGGTGAACGGT 692
QY 259 GGCCTCGGAAGCGATTTTGTGATCGGTCTCTTGAAGAGTTGGGCCCTTGGCCATGCAAT 318
DB 693 GGCCCTGGGTGTTCTTCCCTACCGGTCTCTCATGAGACTTGGCCCTAGCAGCATCAAC 752
QY 319 TCGACTTTTGTGACTATCATCAACCCCTCACTCGTGGAGCAGGTTCTTCCAAATTTACTTC 378
DB 753 AAGAAGATCCAGCGGTCTACATGACTAGCTAGCTTGAAGTCAACGCGTCCGCTGATCTTC 812
```


QY 379 CTGTCACAGCATTGGAGTGGCTTTTCATATAGTGAATACGGTGTGATGGTCCATTAAAC 438
 Db 813 CTGACACAGCCTGTCAATGTCGTTACTCTACAGTAACCTCTGCTGCACGACAGGTC 872
 QY 439 CCTGTACACTGGGTCGTCGA 458
 Db 873 GCTGCTGGCAAGGACGCTCTA 892

RESULT 10

US-08-608-267-1
 ; Sequence 1, Application US/08608267
 ; Patent No. 5688663
 ; GENERAL INFORMATION:
 ; APPLICANT: Yaver, Debbie Sue
 ; APPLICANT: Thompson, Sheryl Ann
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5688663 No. 5688663disk of No. 5688663th America, Inc.
 ; STREET: 405 Lexington Avenue, Suite 6400
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/608,267
 ; FILING DATE: 28-FEB-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/309,341
 ; FILING DATE: 20-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lowney, Karen A.
 ; REGISTRATION NUMBER: 31,274
 ; REFERENCE/DOCKET NUMBER: 4247.000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 867 0123
 ; TELEFAX: 212 867 0298
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2068 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Aspergillus niger
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: 572..632
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: join (571..633)
 US-08-608-267-1

Query Match 4.7%; Score 76; DB 1; Length 2068;
 Best Local Similarity 55.8%; Pred. No. 7.5e-13;
 Matches 145; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 199 TGGTCTTCGAGACGACATACCAAGTGCACCTATACATTCGTGGTGAATGCT 258
 Db 633 TGGTCTTCGAGTCTCGCAATACCAAGTGCACCTATACATTCGTGGTGAATGCT 692
 QY 259 GCCTCTGGAAGCGATCTTCGAGAGAGTTGGGCGCTTGCATGTCAT 318
 Db 693 GCCTCTGGAAGCGATCTTCGAGAGAGTTGGGCGCTTGCATGTCAT 752

QY 319 TCGACTTTTCATGACTACATCAACCCCTACCTCGTGGAGAGGCTCTCCAAATTTACTATTC 378
 Db 753 AAGAAGATCAGCGCGGTCTCAATAGTACGCTTGGAACTCCAAACGGGTCGTGATCTTC 812
 QY 379 CTGTCCAGCAGCATTTGGGAGTGGCTTTTCATATAGTGAATACGGTGTGATGGTCCATTAAAC 438
 Db 813 CTGACACAGCCTGTCAATGTCGTTACTCTACAGTAACCTCTGCTGCACGACAGGTC 872
 QY 439 CCTGTACACTGGGTCGTCGA 458
 Db 873 GCTGCTGGCAAGGACGCTCTA 892

RESULT 11

US-08-608-452-1
 ; Sequence 1, Application US/08608452
 ; Patent No. 5693510
 ; GENERAL INFORMATION:
 ; APPLICANT: Yaver, Debbie Sue
 ; APPLICANT: Thompson, Sheryl Ann
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5693510 No. 5693510disk of No. 5693510th America, Inc.
 ; STREET: 405 Lexington Avenue, Suite 6400
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/608,452
 ; FILING DATE: 28-FEB-1996
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/309,341
 ; FILING DATE: 20-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lowney, Karen A.
 ; REGISTRATION NUMBER: 31,274
 ; REFERENCE/DOCKET NUMBER: 4247.000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 867 0123
 ; TELEFAX: 212 867 0298
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2068 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Aspergillus niger
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: 572..632
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: join (571..633)
 US-08-608-452-1

Query Match 4.7%; Score 76; DB 1; Length 2068;
 Best Local Similarity 55.8%; Pred. No. 7.5e-13;
 Matches 145; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 199 TGGTCTTCGAGACGACATACCAAGTGCACCTATACATTCGTGGTGAATGCT 258
 Db 633 TGGTCTTCGAGTCTCGCAATACCAAGTGCACCTATACATTCGTGGTGAATGCT 692


```
;
;
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 8648.44US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2632 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 696...2291
; OTHER INFORMATION:
;
; US-08-329-892B-32
;
; Query Match 4.1%; Score 66.2; DB 3; Length 2632;
; Best Local Similarity 53.2%; Pred. No. 7.7e-10;
; Matches 140; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
;
; QY 187 CATACCTCTCTCTGGTTCITCGAAGCCGACACATAACCCAGAAACTGCACCTATCACATTG 246
; Db ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
;
; Db 1113 CATTCTCTTTTGGACTTTTGAAGTAGAAGGATCCTCGAAAGGATCGGTCATCCCTT 1172
;
; QY 247 TGGTTGAATGGTGGCCCTGGAAGGATTTGTGATGGTCTCTTCGAAAGAGTTGGGCCCT 306
; Db ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
;
; Db 1173 TGGTTGAACGGGGTCCAGGTTGTTCTTACIACCGGGCTGTCTTGTATTAGGACCC 1232
;
; QY 307 TGCCATGTCATTCGACTTTTGTGATGACTACATCAACCCICACTCGTGGAGACGAGTCTCC 366
; Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; Db 1233 TCATCCATTGGACCTGATTTGAACCCATCGGAACCCCTTACTCTTGGAAACGCAATGCC 1292
;
; QY 367 AATTACTATTCCTGTCGCCAGCCATTGGGAGTCGGCTTTTCATATAGTATACGGTTGAT 426
; Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; Db 1293 ACCGTGATCTTCCTTGACCAACCCGTCAACGTTGGGTCTCGTATCCGGGTCTCAGGT 1352
;
; QY 427 GGGTCCATTAAACCTGTAACTGG 449
; Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; Db 1353 GTTCCACACTGTCGCCGCTGG 1375
```

Search completed: September 16, 2003, 14:45:40
Job time : 150 secs

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 319

; LENGTH: 1814

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-849-626-319

Query Match

Best Local Similarity 3.5%; Score 56.4; DB 10; Length 1814;

Mismatches 0; Pred. No. 1.5e-07;

Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

QY 139 CCGGGTCTCAATCCCTACTCTGGATATGTGACACCTCTCCCGAGTCCCATACCTTCTTC 198

Db 139 CCGTCTTTCCCGCAGTACTCCGGTACCTCAAAAGCTCCGGCTCCAAGCACCTCCCACTAC 198

QY 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTAICACATCTGTGTTGAATGGT 259

Db 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTAICACATCTGTGTTGAATGGT 259

QY 259 GGGCTTGAAGGATTTTGATCGGTCTCTTGAAGAGTGGGCCCTTGCCTATGTCAAT 318

Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTCACAGAGCATGGCCCTTCTTGGTCCAG 318

QY 319 TCGACTTTTATGACT---ACATCAACCTCTACTCGTGAACGAGGTCTCCAAATTTACTA 375

Db 319 CCAGATGGTGTACCTCGGTGAGTACAAACCCCTATTTCTTGAATCTGATGGCAATGTGTA 378

QY 376 TTCTCTGTCCAGCATTTGGGAGTCGGCTTTTCATATAGTAT 417

Db 379 TACTTGGAGTCCCACTGGGTGGGCTTCTCTACTCCGAT 420

RESULT 5

US-09-476-300-319

; Sequence 319, Application US/09476300

; Publication No. US20030125245A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C3

; CURRENT APPLICATION NUMBER: US/09/476,300

; CURRENT FILING DATE: 1999-12-30

; NUMBER OF SEQ ID NOS: 785

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 319

; LENGTH: 1814

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-476-300-319

Query Match

Best Local Similarity 3.5%; Score 56.4; DB 11; Length 1814;

Mismatches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

QY 139 CCGGGTCTCAATCCCTACTCTGGATATGTGACACCTCTCCCGAGTCCCATACCTTCTTC 198

Db 139 CCGTCTTTCCCGCAGTACTCCGGTACCTCAAAAGCTCCGGCTCCAAGCACCTCCCACTAC 198

QY 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACATTTGTTGAATGGT 258

Db 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACATTTGTTGAATGGT 258

QY 259 GGGCTTGAAGGATTTTGATCGGTCTCTTGAAGAGTGGGCCCTTGCCTATGTCAAT 318

Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTCACAGAGCATGGGCCCTTCTTGGTCCAG 318

QY 319 TCGACTTTTATGACT---ACATCAACCTCTACTCGTGAACGAGGTCTCCAAATTTACTA 375

Db 319 CCAGATGGTGTACCTCGGTGAGTACAAACCCCTATTTCTTGAATCTGATGGCAATGTGTA 378

QY 376 TTCTGTCCCAAGCCATTTGGAGTCCGGCTTTTCATATAGTAT 417

Db 379 TACCTGGAGTCCCGAGTGGGGTGGGTTCTCCTACTCCGAT 420

RESULT 6

US-10-113-872-319

; Sequence 319, Application US/10113872

; Publication No. US20030170255A1

; GENERAL INFORMATION:

; APPLICANT: Watanabe, Yoshihiro

; APPLICANT: Hendelerson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Sleath, Paul R.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Gary R.

; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C19

; CURRENT APPLICATION NUMBER: US/10/113.872

; CURRENT FILING DATE: 2002-03-28

; NUMBER OF SEQ ID NOS: 2011

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 319

; LENGTH: 1814

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-113-872-319

Query Match

Best Local Similarity 3.5%; Score 56.4; DB 12; Length 1814;

Mismatches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

QY 139 CCGGGTCTCAATCCCTACTCTGGATATGTGACACCTCTCCCGAGTCCCATACCTTCTTC 198

Db 139 CCGTCTTTCCCGCAGTACTCCGGTACCTCAAAAGCTCCGGCTCCAAGCACCTCCCACTAC 198

QY 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACATTTGTTGAATGGT 258

Db 199 TGGTCTTTCGAAGCCAGACATAACCCAGAACTCCACCTATCACATTTGTTGAATGGT 258

QY 259 GGGCTTGAAGGATTTTGATCGGTCTCTTGAAGAGTGGGCCCTTGCCTATGTCAAT 318

Db 259 GGTCCGGCTGCGAGTCACTAGATGGCTCTCACAGAGCATGGGCCCTTCTTGGTCCAG 318

QY 319 TCGACTTTTATGACT---ACATCAACCTCTACTCGTGAACGAGGTCTCCAAATTTACTA 375

Db 319 CCAGATGGTGTACCTCGGTGAGTACAAACCCCTATTTCTTGAATCTGATGGCAATGTGTA 378

QY 376 TTCTGTCCCAAGCCATTTGGAGTCCGGCTTTTCATATAGTAT 417

Db 379 TACCTGGAGTCCCGAGTGGGGTGGGCTTCTCCTACTCCGAT 420

RESULT 7

US-10-017-754-319

; Sequence 319, Application US/10017754

; Publication No. US20030054363A1

; GENERAL INFORMATION:

; APPLICANT: Henderson, Robert A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Watanabe, Yoshihiro

; APPLICANT: Johnson, Jeffrey C.

; APPLICANT: Retter, Marc W.

; APPLICANT: Marnerakis, Margarita

; APPLICANT: Carter, Darrick

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: McNabb, Andria

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C18

; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319
; LENGTH: 1814
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-754-319

Query Match 3.5%; Score 56.4; DB 14; Length 1814;
Best Local Similarity 52.5%; Pred. No. 1.5e-07;
Matches 148; Conservative 0; Mismatches 131; Indels 3; Gaps 1;
QY 139 CGGGGTGCAATCTTACCTGATGATGTCGACACCTCTCCCGAGTCCCATACCTTCTTC 198
DB 139 CGGTCTTTCCGCGAGTACCTCGGCTACCTCAAAAGCTCCGGCTCCCAAGCACCCTCCACTAC 198
QY 199 TGGTTCTCGAAGCCAGACATACCCAGAACTGCACCTATACATTTGGTGGTGAATGGT 258
DB 199 TGGTTCTCGAGTCCAGAGGATCCCGAGACAGCCCTGTGGTCTTGGCTCAATGGG 258
QY 259 GGGCCGGAAGGAGTCTTTTGATGCGGTCTCTGCGAAGAGTTGGGCGCTTGGCAATGCAAT 318
DB 259 GGTCCGGGTGCGAGTCACTAGATGGGTCTCTCAGACAGCATGGCCCTCTCTGGTCCAG 318
QY 319 TCGACTTTTGATGACT--ACATCAACCTCCTGCTGGAACGAGTCTCCCAATTTACTA 375
DB 319 CCAGATGGGTGCACCTGGAGTACAAACCCCTATCTTGGAACTGATGCAATGCTGTA 378
QY 376 TTCTCTGCCAGCATTGGGAGTCGGCTTTTCATATAGTAT 417
DB 379 TACCTGGAGTCCCGAGCTGGGTGGCTTCTCTACTCCGAT 420

RESULT 9
US-09-925-302-331
; Sequence 331, Application US/09925302
; Patent No. US2002044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 331
; LENGTH: 2864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2850)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2858)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2860)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-331

Query Match 3.4%; Score 54.8; DB 9; Length 2864;
Best Local Similarity 52.1%; Pred. No. 6.6e-07;
Matches 147; Conservative 0; Mismatches 132; Indels 3; Gaps 1;
QY 139 CCGGTGTCAAAATCTTCTGATATGTCGACACCTCTCCCGAGTCCCATACCTTCTTC 198
DB 171 CCGTCTTTCCGCGAGTACTCCGGTACCTCAAAAGCTCCGGCTCCAGCACCCTCCACTAC 230
QY 199 TGGTTCTCGAAGCCAGACATACCCAGAACTGCACCTATACATTTGGTGGTGAATGGT 258
DB 231 TGGTTTGGAGTCCCGAGGATCCCGAGACACCCCTGTGGTCTTGGCTCAATGGG 290
QY 259 GGGCCCTGGAAGCGATTTCTTGTATCGSTCTCTTTCGAGAGTTGGGCGCTTGGCATGTCAT 318
DB 291 GGTCCCGCTGCAGCTCACTAGATGGGTCTCTCAGACAGCATGGCCCTCTCTGGTCCAG 350
QY 319 TCGACTTTTGATGACT--ACATCAACCTCCTGCTGGAACGAGTCTCCCAATTTACTA 375
DB 351 CCAGATGGGTGTCACCCCTGGAGTACAAACCCCTATCTTGGAACTGATGCAATGCTGTA 410
QY 376 TTCTCTGCCAGCATTGGGAGTCGGCTTTTCATATAGTAT 417
DB 411 TACCTGGAGTCCCGAGCTGGGTGGCTTCTCTACTCCGAT 452

RESULT 10
US-09-790-399-7

Sequence 7, Application US/09790399
 Patent No. US20020038000A1
 GENERAL INFORMATION:
 APPLICANT: Gold, Larry
 APPLICANT: Tuerk, Craig
 APPLICANT: Priboj, David
 APPLICANT: Smith, Jonathan D.
 TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
 FILE REFERENCE: NEX02/C1-CON2
 CURRENT APPLICATION NUMBER: US/09790,399
 CURRENT FILING DATE: 2001-02-22
 PRIOR APPLICATION NUMBER: 09/197,649
 PRIOR FILING DATE: 1998-11-23
 PRIOR APPLICATION NUMBER: 07/829,461
 PRIOR FILING DATE: 1992-01-31
 PRIOR APPLICATION NUMBER: 07/739,055
 PRIOR FILING DATE: 1991-08-01
 PRIOR APPLICATION NUMBER: 07/561,968
 PRIOR FILING DATE: 1990-08-02
 NUMBER OF SEQ ID NOS: 26
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 7
 LENGTH: 390
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Sequence
 OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
 OTHER INFORMATION: fragments having NcoI restriction sites.
 US-09-790-399-7

Query Match 3.0%; Score 47.6; DB 9; Length 390;
 Best Local Similarity 45.3%; Pred. No. 5.4e-05;
 Matches 173; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY	1074	GGCAAGGACTCTGTCATGACGCTATCGCGCTCAACATCACTACACCGAGTCCCAATAA	1133
Db	2	GGCCATGGAGCGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	61
QY	1134	TGACGCTACTACGCTTCCAGCAAAACAGCGGACTTGTCTGGCCCAACTTCATCGAGA	1193
Db	62	CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	121
QY	1194	CCTCGAGGAGATCTTCTCTCCCGTGGTGTCTCCCTCATCTATGGGACGCGGATTA	1253
Db	122	CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	181
QY	1254	CATCTCAACTGTTTCGGCGGTGAGCGGTTTCCCTCGTGGCAACTACTTCCCAAGCGCG	1313
Db	182	CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	241
QY	1314	CCAGTTCCGAAGCGGAGGTACACGCCCTGAAAGTCAACGGGTGAGTATGGGAAAC	1373
Db	242	CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	301
QY	1374	TCGCGAGTAGTAAATTTCTCTTCACTCGCTCTATGAGGAGGCGCATGAAGTCCCAATA	1433
Db	302	CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	361
QY	1434	CTACAGCCCATCGCTCCCTG	1455
Db	362	CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	383

RESULT 11

US-10-084-018-4
 Sequence 4, Application US/10084018
 Publication No. US20020160499A1
 GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 Hawkins, Phillip R.
 Hillman, Jennifer L.
 Lal, Preeti

Goli, Surya K.
 TITLE OF INVENTION: NOVEL HUMAN SERINE
 CARBOXYPEPTIDASE
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/084,018
 FILING DATE: 25-Feb-2002
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/299,689A
 FILING DATE: <Unknown>
 APPLICATION NUMBER: 08/828,488
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0241 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1551 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: MMLR3DT01
 CLONE: 566993
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-10-084-018-4

Query Match 3.0%; Score 47.6; DB 13; Length 1551;
 Best Local Similarity 46.9%; Pred. No. 0.00012;
 Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY	166	GTGACACCTCTCCGAGTCCCATACCTTCTTCGTGTTCTTCGAGCCGACACATAACCA	225
Db	320	GTGAATAAGACTTACACAGCAACCTCTTCTGTGTTCTCCAGCTCAGATACAGCCA	379
QY	226	GAAACTGCACCTATCACATTTGGTTGAATGGTGGCCCTGAGCGGATTTCTTTGATCGGT	285
Db	380	GAGAATGCCCACTAGTTCTCTGCTACAGGTGGGCGGGAGGTTCATCATGTTWGA	439
QY	286	CTCTTCGAGAGTTGGCCCTTCGCTGCTCAATTCGACTTTTGATGACTACATCAACCT	345
Db	440	CTCTTTGTGGAACATGGGCTTATGTTGTACAAAGTAACATGACCTTGCCTGACAGAC	499
QY	346	CACCTCGTGGACGAGGCTTCCCAATTTACTATTTCTCTCCAGCATTTGGAGTCGCTTT	405
Db	500	TTCCCCCTGGACCAACAGCTCTCCATGCTTTACATTCAGAAATCCAGTGGSCACAGCTTC	559
QY	406	TCATATAGTGATACGGTTTGATGGTCCATTAACTTACCTTGGGTCGCTGCGAAATTCG	465
Db	560	AGTTTACTCATGATACCCACCGAATGAGTCAATGAGGACGATGTAGCACGGGATTTA	619
QY	466	AGCTTTGCAGGAGTTTCAAG	483
Db	620	TACAGTGCATTAATTCAG	637

RESULT 12

```
US-09-729-674-109
; Sequence 109, Application US/09729674
; Patent No. US20010033335A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Meirberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6055-64X
; CURRENT APPLICATION NUMBER: US/09/729,674
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 109
; LENGTH: 1684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-729-674-109

Query Match      3.0%; Score 47.6; DB 9; Length 1684;
Best Local Similarity 46.9%; Pred. No. 0.00013;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 166 GTCGACACCTCTCCGAGTCCCAATCTCTTCGTTCTTCGAGCCAGACATACACCA 225
Db 332 GTGATAGACTTACACAGCAACCTCTCTTCGTTCTTCGAGCCAGACATACACCA 391
QY 226 GAAACTGACATATACATGTTGTTGAATGGTGGCCCTGGAACGGATCTTTGATCGGT 285
Db 392 GAAGATGCCCGAGTAGTCTCTGGCTACAGGTGGCGGGAGGTTCATCATCTTTGGA 451
QY 286 CTCCTGCAAGATGGGCCCTTCGATCTCAATTCGACTTTTGTATGACTACATCAACCT 345
Db 452 CTCCTTGGCAATAGTGGCCCTATGTTGTCACAGTACATGACCTTCGTCAGAGAC 511
QY 346 CACTGCTGGACAGAGTCTCCAAATTTACTATTCCTGTCGCCAGCCATGGGAGTGGCTTT 405
Db 512 TTCCCTGGACACACAGCTCTCCATGCTTTACATTTGACATCCAGTGGGCACAGCTTC 571
QY 406 TCATATAGTATAGGTTGATGGTTCATTAACCTGTAACTGGGTGCTCGAAATTCG 465
Db 572 AGTTTACTGATATACCCAGGATATGACAGTCAATGAGGAGTGTAGACAGGGATTTA 631
QY 466 AGCTTTGAGGAGTTTCAG 483
Db 632 TACATGCACTAATTCAG 649

RESULT 13
US-09-796-753-39
; Sequence 39, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30

Query Match      3.0%; Score 47.6; DB 11; Length 1697;
Best Local Similarity 46.9%; Pred. No. 0.00013;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 166 GTCGACACCTCTCCGAGTCCCAATCTCTTCGTTCTTCGAGCCAGACATACCA 225
Db 338 GTCAATAAGACTTACACAGCAACCTCTCTTCGTTCTTCGAGCCAGACATACCA 397
QY 226 GAAACTGACCTATACATGTTGTTGAATGGTGGCCCTGGAACGGATCTTTGATCGGT 285
Db 398 GAAGATGCCCGAGTAGTCTCTGGCTACAGGTGGCGGGAGGTTCATCCATGTTTGA 457
QY 286 CTCCTGCAAGATGGGCCCTTCGATCTCAATTCGACTTTTGTATGACTACATCAACCT 345
```

```

; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-909-320-163

Query Match      3.0%; Score 47.6; DB 10; Length 2076;
Best Local Similarity 46.9%; Pred. No. 0.00014;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 166 GTGCACACCTCTCCCGAGTCCCAACCTCTCTCTGGTTCTTCGGAAGCCAGACATAAOCOA 225
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 334 GTCAATAAGACTTTACAACAGCACCTCTCTCTGGTTCTTCACAGCTCAGATACAGCCA 393
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 226 GAAACTGCACCTATCACATTTGTTGCAATGGTGGCCCTGGAACGGATCTTTTCATCGGT 285
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 GAGATGCCCCAGTAGTTCTTCGCTACAGGGTGGCCGGGAGGTTTCATCCATGTTTGG 453
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 286 CTCTCGAAGAGHTGGGCCCTTGCCATGTCAAITCGACITTTGATGATACATCAACCT 345
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 454 CTCTTTGTGAACATGGGCCCTTATGTGTCACAGTAACATGACCTTGGCGTGACAGAC 513
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 346 CACTCTGGGAACAGGTCTCAATTACTATTCTGTCCGAGCCATTGGGAGTGGGCTTT 405
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 514 TTCCCTCTGGACCACAAGCCTCCATGCTTTACATTGACAATCCAGTGGGCACAGGCTTC 573
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 406 TCATATAGTATACGGTTGTATGGTCCATTAACTCCCTGTAATCGGGTCTGCAAAATTCG 465
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 574 AGTTTACTGATATACCCAGGATATGCAATGAGGACGATGATGACACGGGATTTA 633
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 466 AGCTTGGCAGGATTCAG 483
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 634 TACAGTGCATAATTCAG 651
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-909-320-163

Query Match      3.0%  Score 47.6; DB 10; Length 2076;
Best Local Similarity 46.9%  Pred. No. 0.00014;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY      166  GTGCACACCTCTCCCGAGTCCCATACCTTCTCTGGTCTCTTCGAGCCAGACATACACCA 225
Db      166  GTGAATGAGACTTTACACAGCAACCTCTCTCTGGTCTCTTCACAGCTCAGATACAGCCA 393

QY      226  GAAACTGCACCTATCACATTTGTTGAATGGTGGCCCTCGAAGCGGATTCTTTTCATCGGT 285
Db      226  GAAAGATGCCCCAGTAGTTCTCTGGCTACAGGGTGGCCGGAGGTTTCATCCATGTTTGA 453

QY      286  CTCTCTCGAAGAGTGTGGCCCTTGCCTATGTCAAATTCGACITTTGAAGTACATCAACCCCT 345
Db      286  CTCTTTGTGGAACATGGGCCCTTATGTGTGCACAGTAACATGACCTTGGGTGACAGAGAC 513

QY      346  CACTCTGTGGAACGAGGTCCTCCAAATTTACTATTCCTGTCGCCAGCCATTGGGAGTCGGCTTT 405
Db      346  TTCCCTGTGGACCAACAGCGTCTCCATGCTTTACATTGACAACTCCAGTGGGCACAGGCTTC 573

QY      406  TCATATAGTAGTACGGTTGATGGTCCATTAACCCCTGTAACCTGGTGGGTCGTCGAAATTCG 465
Db      406  AGTTTTACTAGTATCATACCCACGGATATGCACTCAATGAGCAGCATGTAGCACGGGATTTA 633

QY      466  AGCTTTGTCAGGAGTTCAG 483
Db      466  TACAGTGCACATAATTCAG 651

RESULT 15
US-09-909-088B-163
; Sequence 163, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hauspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.

```

APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/909,088B
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 163
LENGTH: 2076
TYPE: DNA
ORGANISM: Homo sapiens
US-09-909-088B-163

Query Match 3.0%; Score 47.6; DB 10; Length 2076;
Best Local Similarity 46.9%; Pred. No. 0.00014;
Matches 149; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 166 GTCGACACCTCTCCCGAGTCCCATACCTCTCTCTGCTTCTCGAAGCCACACATAACCCCA 225
DB 334 GTGAATAAGACTTACACACACACCTCTCTCTGCTTCTCGAAGCCACACATAACCCCA 393

QY 226 GAACTGACCTATACATGCTGTTGAATGGTGGCCCTGGGAAGCCGATCTTTGATCGGT 285
DB 394 GAAGATGCCCGAGTAGTCTCTGCTACAGGTGGGCGCGGAGGTTCATCCATGTTTGA 453

QY 286 CTCTTCGAGAGAGTTGGCCCTTCCCATGCAATTCGACTTTGATGACTACATCAACCCCT 345
DB 454 CTCTTTGTGAACATGGCCCTTATGTTGTCAACAGTAAGACCTTGGTGACAGAGAC 513

QY 346 CACTCGTGGAGAGGTCTCCAAATTTACTATTCTCTGTCAGCCATTTGGGAGTGGCTTT 405
DB 514 TTCCCTGGACCAACAGCTCTCCATGCTTTACATTTGACAATCCAGTGGGCACAGGCTTC 573

QY 406 TCATATAGTAGTACGGTTGATGGTCCATTAACCCCTGTAACCTGTCGTCGAAATTCG 465

DB 574 AGTTTACTGATGATACCCACGGATATGCGATGCAATAGGAGCAGATGTAGCAGGGATTTA 633
QY 466 AGCTTTGCGAGGAGTTTCAG 483
DB 634 TACAGTGCCTAATTCAG 651

Search completed: September 16, 2003, 18:38:56
Job time : 1478 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 14:42:53 ; Search time 4185 Seconds
(without alignments)
9338.494 Million cell updates/sec

Title: us-09-712-338-1_COPY_55_1662

Perfect score: 1608

Sequence: 1 cttccaggaagtacacggc.....gcattgcagtggtgtatg 1608

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	138.4	8.6	588	9	AW672518
2	116	7.2	721	14	CA747625
3	92.6	5.8	498	12	BM868983
4	86.6	5.4	963	29	CNS0625E

5	86.6	5.4	1049	29	CNS06XWN
6	83.6	5.2	696	13	B0634841
7	80.6	5.0	485	12	B1200653
8	77.6	4.8	445	12	B1187544
9	77.6	4.8	475	9	AW677015
10	77	4.8	1054	29	CNS06NN7
11	74	4.6	729	13	B0110016
12	72.8	4.5	1002	29	CNS06ZOK
13	72.4	4.5	662	14	CD044859
14	72.2	4.5	682	14	CD054523
15	70.2	4.4	907	14	CD456151
16	69.2	4.3	675	14	CB937789
17	68.2	4.2	1109	14	CD508724
18	68	4.2	580	13	BW197413
19	67.6	4.2	700	12	B1749802
20	67.4	4.2	685	13	BW194502
21	66.2	4.1	457	14	CD039988
22	66.2	4.1	544	14	CD042968
23	66.2	4.1	712	14	CD045123
24	66	4.1	648	10	BE777028
25	66	4.1	653	14	CD258740
26	65.8	4.1	482	9	AV985320
27	65.8	4.1	550	9	AV957236
28	65.8	4.1	566	13	BW276795
29	65.8	4.1	686	13	BW260398
30	65.4	4.1	287	14	W06667
31	64.8	4.0	590	13	BW198196
32	64.2	4.0	551	9	AV976226
33	64.2	4.0	555	9	AV974977
34	64.2	4.0	560	12	BP007496
35	64.2	4.0	585	9	AV965651
36	64.2	4.0	638	13	BW282449
37	64.2	4.0	652	9	AV991679
38	64.2	4.0	656	13	BW276796
39	64.2	4.0	656	13	BW292110
40	64.2	4.0	675	13	BW274166
41	64.2	4.0	675	13	BW302983
42	64.2	4.0	680	13	BW208298
43	64.2	4.0	684	13	BW248944
44	64.2	4.0	684	13	BW251033
45	64.2	4.0	686	13	BW252952

ALIGNMENTS

RESULT 1
AW672518
LOCUS LG1_360_F03.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
DEFINITION LG1_360_F03.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
EST. sequence.
ACCESSION AW672518 GI:7536439
VERSION AW672518.1
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 588)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Contact: Ebole DJ
Department of Plant Pathology & Microbiology
Texas A&M University
Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
Tel: 979 845 4831
Fax: 979 845 6483
Email: d-ebbole@tamu.edu

Chromatogram file of this sequence is available, see contact person
;Best nr hit (April. 22, 2003) gb|EAA29397.1| hypothetical protein
[Neurospora crassa] 132 2e-30
PCR Primers
FORWARD: T3 primer
BACKWARD: T7 primer
Plate: mgns002 row: M column: 24
seq primer: t3.

```

FEATURES
source
    Location/Qualifiers
        seq_id=1;
        .498
        /organism="Magnaporthe grisea"
        /mol_type="mRNA"
        /strain="Guyll"
        /db_xref="taxon:148305"
        /clone="mgns002xm24"
        /sex="Matl-2 hermaphrodite"
        /cell_type="mycelium"

```

```

/clone_lib="Magnetaporthe grisea NS Uni-zap XR Library"
/notes="Vector: pBluescriptSK-; Site_1: EcoRI; Site_2: XhoI
; Unidirectional cloning. EcoRI site has T3 primer and
predominantly 5' reads. T7 primer on XhoI side of insert.
Nitrogen starvation library. Cells were inoculated into
minimal medium and grown for two days with shaking (150
rpm) at room temperature. Culture was harvested, blended,
inoculated into minimal medium as above for 24 h. Cells
were harvested, washed with water and inoculated into
minimal medium base lacking nitrogen source for 6 h.
Sequences were processed by one of two methods. Where a
full-length alignment to the M. grisea genome sequence was
available, the EST sequence was trimmed according to the
alignment, otherwise sequence quality was assessed using
puredPhrap version 99i019 and trimmed according to phd
files (.0.05) and trimmed for vector seqs."
94 a 186 c 143 g 75 t

```

BASE COUNT	94 a	186 c	143 g	75 t	ORIGIN
Query Match	5.8%;	Score 92.6;	DB 12;	Length 498;	
Best Local Similarity	52.1%;	Pred. No. 8.8e-16;			
Matches 231;	Conservative 0;	Mismatches 209;	Indels 3;	Gaps 1;	
QY	1087	GTCTGGAGGCTATGGCGTCAACATCAACTACCCAGTCCTCAATAATGACGCTCTACIAC	1146		
Db	38	GTCCAGCTGGCGCTGGGCTACCCCTCACTGGACCAGCTCTCTGTCGGCATCGTCAACG	97		
QY	1147	GCTTTCCAGCAACAGGGCAGTTTGCTGGCCCACTTCATCGAAGACCTCGAGGAGATC	1206		
Db	98	GCCTTCCGGTCCATCGGGGACTACAAACGCCCGCGGTGGCTCGAAGCCAATGGCAACCTC	157		
QY	1207	CTTGCTCTCCCGTGGCTGTCTCCCTCATCTATGGCGACGCCGATTCACATCTGCCACTGG	1266		
Db	158	CTCGACAGGGCAAGAGTGCAGCTCATGTACGGCGACCGGATTTCCGCTTCCAACTGG	217		
QY	1267	TTGCGCGGTTCAGCGCGTTTCCCTCGCTGCGAACTACTCCAAAGCCCGCAGTTCGGAAGC	1326		
Db	218	ATGGCGGTTGAAGCGCGCTCCTTGGCCATCCCGTGGAGGAACCGAAGAAAGTTTCGCCGAG	277		
QY	1327	GCAGGGTACACGCCCTGAAATGCAACGG---CGTGCAGTATGGGGAACACTCGCGASTAT	1383		
Db	278	GCGGGCTACAGCCCTCGGGACCAACTGCACCTAGCAGCGCGGTTGGTGGCGCACTAC	337		
QY	1384	GGTAATTTCTCCTTCACCTGGCTCTATGAGGCGAGGCCATGAAGTCCTCACTACCAACCC	1443		
Db	338	TCCAAACCTGACCTTTTGCCCGGCTCTCCAGGCGGGACAGCCGCCACCTTCGTACCAAGCAG	397		
QY	1444	ATCGCCTCCCTGCANTTGTTTAACCGGACTATCTTCGTTGGGATATCCAGAGGGCCAG	1503		

398 CAGACGGCCCTACCGCATCTTCAACCGCGGCCCTGTGTCACAGGACGTGGCGGGCGGCTG 457

1504 AAGAAGATCTGGCCAGCTACAA 1526

458 GTGACACGCGGACGAACCCCGA 480

CNS0625E 963 bp DNA linear GSS 07-JUL-2001
T3 end of clone AY0AA015G04 of library AY0AA from strain CBS 6340
of Kluyveromyces thermotolerans, genomic survey sequence.

ACCESSION AL421896
VERSION AL421896
KEYWORDS GI:12205091
SOURCE GSS.
ORGANISM Kluyveromyces thermotolerans
Eukaryota; Fungi; Ascomycota: Saccharomycetia; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE 1 (bases 1 to 963)
AUTHORS Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Boilotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

REFERENCE 2 (bases 1 to 963)
MEDLINE 20584711
PUBMED 11152876
AUTHORS Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 10. Kluyveromyces thermotolerans
FEBS Lett. 487 (1), 61-65 (2000)

REFERENCE 3 (bases 1 to 963)
MEDLINE 11152885
PUBMED
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES
Source Location/Qualifiers
1..963
/organism="Kluyveromyces thermotolerans"
/mol_type="genomic DNA"
/strain="CBS 6340"
/db_xref="taxon:4916"
/clone="AY0AA015G04"
/clone_lib="AY0AA"
/note="end : T3"
misc_feature <270..>962
/note="similar to Saccharomyces cerevisiae ORF YBR139w [strong similarity to carboxypeptidase]"
evidence=not_experimental

BASE COUNT 242 a 231 c 217 g 272 t 1 others
ORIGIN

Query Match 5.4%; Score 86.6; DB 29; Length 963;
Best Local Similarity 55.5%; Pred. No. 7.7e-14;

Fax: 405 325 7762
Email: broe@ou.edu
Contact Dr. Marian Beremand regarding clone availability. Included is the best homology from a blastx search of Genbank nr_04-09-01 466 2e-46 gi|461830|sp|P34946| CARBOXYPEPTIDASE Sipiir|I338953

Contact Dr. Marian Beremand regarding clone availability Included
is the best homolog from a blastx search of Genbank nr 04-09-01
386 5e-37 g1|461830|sp|P34946| CARBOXYPEPTIDASE Slpir||S38953
carboxypept
Seq primer: T3

```

FEATURES
  source
    High quality sequence stop: 413.
    Location/Qualifiers
      1..445
        /organism="Fusarium sporotrichioides"
        /mol_type="mRNA"
        /strain="Tri 10"
        /db_xref="taxon:5514"
        /clone="a4e08fs"
        /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
        cDNA library"
        /note="Vector: pBlueScript SK-; Site_1: EcoRI; Site_2:
        XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
        ; 3' end of cDNA cloned into XhoI site of pBluescript."
BASE COUNT      113 a      117 c      109 g      106 t
ORIGIN
  Query Match      4.8%; Score 77.6; DB 12; Length 445;
  Best Local Similarity 58.8%; Pred. No. 2.1e-11;
  Matches 153; Conservative 0; Mismatches 104; Indels 3; Gaps 1;
QY      121 GGGCTCTGGAGACTACCCGGGTGTCAAATCCTACTCTGGATATGTCGACACCTCTCC 180
Db      186 GGAATATGGAGACTACCGCAGGTGTCGCCAGCACTCTGGTATTTCTCTGTCGGGAC 245
QY      181 GAGTCCTACCTCTCTCTGGTCTTCGAAAGCCAGACATAAACCCAGAAACTGCACCTATC 240
Db      246 AACATGAACATGTGTCTTGGTCTTCGAGTCCGCCAAGAACGCCAAAACAGCTCCACTG 305
QY      241 ACATTCGTGTGAATGTCGCCCTGGAAGGATCTTTGATCGTCTTCGAGAGTTG 300
Db      306 GGCCTTTGGCTCAATGTCGGCCAGGCTGTAGTTCATGATCGTCTTCCAGGAAAT 365
QY      301 GGCCCTTGGCATGTCAATTGCG---ACTTTTGATGACTACATCAACCCCTCACTCGTGGAA 357
Db      366 GGCCCTTGACCTTCACAGAGGAGGTCACAAACCACTTAACCCCTACTCATGGAAC 425
QY      358 GAGGTCTCCAATTACTATT 377
Db      426 ACATTCGCCAACAATGTATT 445

RESULT 9
AW677015      475 bp      mRNA      linear      EST 19-JUL-2000
LOCUS      DGI_3_C02.bi_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
DEFINITION
  sequence.
ACCESSION      AW677015.1 GI:7550690
VERSION
KEYWORDS
SOURCE      Sorghum bicolor (sorghum)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
  clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
  1 (bases 1 to 475)
  Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
  ,L.H.
  An EST database from Sorghum: dark-grown seedlings
  Unpublished
  Contact: Cordonnier-Pratt MM
  Laboratory for Genomics and Bioinformatics
  The University of Georgia, Department of Plant Biology
  Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
  Tel: 706 542 1860
  Fax: 706 583 0210
  Email: mmpratt@uga.edu
  Sequences have been trimmed to exclude PolyA, vector and regions
  below Phred quality 16. The threshold for highest quality sequence
  is 20.
  Seq primer: JEN REV
  High quality sequence stop: 370
  POLYA=No.
  Location/Qualifiers

FEATURES
  source
    1..475
      /organism="Sorghum bicolor"
      /mol_type="mRNA"
      /db_xref="taxon:4558"
      /clone_lib="Dark Grown 1 (DGI)"
      /note="Organ: 5-day-old dark-grown seedlings; Vector:
      Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
      made from poly-A RNA in the cloning vector lambda ZAP II.
      Clones to be sequenced were prepared by mass excision."
BASE COUNT      107 a      129 c      124 g      115 t
ORIGIN
  Query Match      4.8%; Score 77.6; DB 9; Length 475;
  Best Local Similarity 62.2%; Pred. No. 2.2e-11;
  Matches 122; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY      120 GGGCTCTGGAGACTACCCGGGTGTCAAATCCTACTCTGGATATGTCGACACCTCTCC 179
Db      208 GGGCATCTGTGAGACCACTCCAGGGGTTAATCATCTACTCGGGGTATCTCTCCGTGGGCTC 267
QY      180 CGAGTCCCATACCTCTCTCTGGTCTTCGAAAGCCAGACATAACCCAGAAACTGCACCTAT 239
Db      268 TATATGAACATGTGTCTTGGTCTTCGAGGCCGCCAACACACCCAGCAAGCTCCCT 327
QY      240 CACATCTGTGTGAATGTCGCCCTGGAAGCGATCTTTGATCGTCTTCGGAAGATT 299
Db      328 GGCTGCTGTTCACGGCGGCCCTGGCTGCTCTCTCCATCATCGTCTGTTCAGAGAA 387
QY      300 GGCCCTTGGCATGTC 315
Db      388 TGGTCTCTGCCACTTC 403

RESULT 10
CNS06NN7      1054 bp      DNA      linear      GSS 04-JUL-2001
LOCUS      T7 end of clone AU00A015E08 of library AU00A from strain CBS 3082
DEFINITION
  of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION      AL406985
VERSION      AL406985.1 GI:12171743
KEYWORDS
SOURCE      Saccharomyces kluyveri
ORGANISM
  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
  1 (bases 1 to 1054)
  Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
  Boitot,F., Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
  de-Montigny,J., DuJon,B., Durrens,P., Lepingle,A., Liorente,B.,
  Malpertuy,A., Neveuglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
  Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
  Wincker,P. and Weissbach,J.
  Genomic exploration of the hemiascomycetous yeasts: 1. A set of
  yeast species for molecular evolution studies
  JOURNAL      FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE      20584711
  PUBMED      11152876
REFERENCE
  2 (bases 1 to 1054)
  Neveuglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
  Gaillardin,C. and Casaregola,S.
  Genomic exploration of the hemiascomycetous yeasts: 9.
  Saccharomyces kluyveri
  JOURNAL      FEBS Lett. 487 (1), 56-60 (2000)
  MEDLINE      20584719
  PUBMED      11152884
REFERENCE
  3 (bases 1 to 1054)
  Genoscope.
  Direct Submission
  Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
  2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
  seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
  This GSS is part of a random genomic sequencing program of thirteen
  yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

```



```

Kluyveromyces marxianus var. marxianus
FEBS Lett. 487 (1), 71-75 (2000)
JOURNAL MEDLINE
20584722
PUBMED 11152887
REFERENCE 3 (bases 1 to 1002)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    source
    1..1002
        /organism="Kluyveromyces marxianus"
        /mol_type="genomic DNA"
        /strain="CBS 712"
        /variety="marxianus"
        /db_xref="taxon:4911"
        /clone_lib="AZ0AA003B09"
        /clone="AZ0AA"
        /note="end : T3"
    misc_feature
        complement(<8. .>994)
        /note="similar to Saccharomyces cerevisiae ORF YMR297w [
        PR1 ; carboxypeptidase Y, serine-type protease ]"
    misc_feature
        complement(<11. .>778)
        /note="similar to Saccharomyces cerevisiae ORF YBR139w [
        strong similarity to carboxypeptidase ]"
        /evidence=not_experimental
    BASE COUNT 267 a 259 c 240 g 235 t
    ORIGIN
        Query Match 4.5%; Score 72.8; DB 29; Length 1002;
        Best Local Similarity 53.5%; Pred. No. 9e-10;
        Matches 152; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
    QY 187 CATACCTCTCTCGGTCTTCGAGCCAGACATACCCAGAACTGCACCTATCACATTG 246
        || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
    Db 673 CACTTCTTCTACTGGTCTTTGAGTCCAGAAACGCCCCAGAACGACCCCTGTATCTTG 614
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 247 TGGTTGAATGGTGGCCCTGGAACCGAATCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCT 306
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 613 TGGTTGAACGGTGGCCAGGGTTCCTCTTTGACTGGTGTCTTCGAATTTGGGTCCA 554
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 307 TGGCATGTCAATTCGACTTTTGATGACTACATCAACCTTCACCTCGTGGACGAGGCTCC 366
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 553 TCTTCATTTGGCGAGAGGTGAAGCCAAATTTACACCCACACTCTTTGGAACAAGCCT 494
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 367 AATTTACTATCTCTGCCAGCCATTGGGAGTCCGGCTTTTCATATAGTATACGGTTGAI 426
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 493 TCCGTTATCTCTTTGACACGACGACGTCACGTTGGTTACTCTTACTCTTCATCAGAAGT 434
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 427 GGGTGCATTAACCTCTACTGCGGGTCGTCGAAATTCGAGCTT 470
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 433 GTCTCAACACTGTTCGTCAGGTGAAGGACGTGTATCGGTTCTT 390
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    RESULT 13
    CD044859
    LOCUS psHB016xG08f.183244 psHB: Infected hypocotyl soybean host. 48 hrs
    DEFINITION post infection Phytophthora sojae cDNA clone SHB016G08 5, mRNA
    sequence.
    CD044859
    Kluyveromyces marxianus var. marxianus
    GI:30498452
    EST.
    SOURCE Phytophthora sojae
    ORGANISM Phytophthora sojae
    Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
    Phytophthora.
    1 (bases 1 to 662)
    REFERENCE Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
    USDA-IFARS: Expression of Phytophthora sojae genes during infection
    and propagation
    JOURNAL Unpublished
    COMMENT Contact: Tyler B
    Tyler lab
    VBI
    1880 Pratt Dr., Blacksburg, VA 24061, USA
    Tel: 540-231-7318
    Email: bmtyer@vt.edu
    PCR Primers
    FORWARD: BK reverse
    Plate: 016 row: G column: 08
    Seq primer: BK reverse
    High quality sequence stop: 662.
    FEATURES
    source
    1..662
        /organism="Phytophthora sojae"
        /mol_type="mRNA"
        /db_xref="taxon:67593"
        /clone="SHB016G08"
        /tissue_type="infected host tissue"
        /cell_line="P6497"
        /dev_stage="48 hour post infection"
        /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
        post infection"
        /note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI;
        USDA-IFARS: Expression of Phytophthora sojae genes during
        infection and propagation."
    BASE COUNT 130 a 204 c 208 g 120 t
    ORIGIN
        Query Match 4.5%; Score 72.4; DB 14; Length 662;
        Best Local Similarity 53.0%; Pred. No. 9.2e-10;
        Matches 179; Conservative 0; Mismatches 156; Indels 3; Gaps 1;
    QY 88 GTCACCATCGGTACAAGAACCGGGGACAGAGGCGCTCTCGAGACTACACCCGGGTGC 147
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 217 GTCACTAACAGTCTGAGACACCGACACAGATTCTCTCGGGGATACGAATCAGAG 276
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 148 AATTCCTACTCTGGATATGCGACACCTCTCCCGAGTCCCATACCTCTCTCTGTTCTTC 207
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 277 ACGGGCTTACATTAAT---TACCCACAAGSACGACGACCACTACTTCTACTGTGGTGC 333
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 208 GAAGCCAGACATACCCAGAACTGCACCTATCATATTGGTTGAATGGTGGCCCTGGA 267
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 334 GAGTCGGCGACAGCTCGCGAGAGGACCCCTTGGTGTCTGGCTCAGGGGCCCCCGGC 393
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 268 AGCGATTCTTTGATCGGTCTCTTGAAGAGTTGGGCCCTTGGCCATGTCAATTCGACTTTT 327
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 394 TGCTCCAGCATGATGGCGCTACTTGGCCGAGAAATGGTCTCTGCCACGTCGACCCGACCTG 453
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 328 GATGACTATACATCAACCCCTACTCGTGGAAAGAGGTCTCCAAATTTACTATTCCTGTC 387
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 454 TCCACCAAGACCAACCGCTACTCGTGGAAAGAGGACGACCAATGTTATTGGTGGACCAA 513
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY 388 CCATTGGGAGTCGGCTTTTCATATAGTGATACGTTGA 425
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db 514 CCCACAGGCGTGGCTACTCGTACGGCCCCCAAGGTGA 551
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    RESULT 14
    CD054523
    LOCUS CD054523 682 bp mRNA linear EST 05-JUN-2003
    DEFINITION HO01B22r HO Hordeum vulgare cDNA clone HO01B22 5-PRIME, mRNA
    sequence.

```

DEFINITION	FEATURES	BASE COUNT	ORIGIN
Fg03_10g07_R Fg03 AAFRC_ECORC_Fusarium_graminearum_mycelium_trichotheceae_product ion Gibberella zeae cDNA clone Fg03_10g07, mRNA sequence.	<p>ACCESSION CD054523.1 GI:30595544</p> <p>VERSION EST.</p> <p>KEYWORDS</p> <p>SOURCE Hordeum vulgare</p> <p>ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.</p> <p>REFERENCE 1 (bases 1 to 682)</p> <p>AUTHORS Zierold,U. and Schweizer,P.</p> <p>TITLE Barley ESTs from pathogen-attacked leaf epidermis</p> <p>COMMENT Unpublished</p> <p>Contact: Patrick Schweizer</p> <p>Transcriptome Analysis, Cytoogenetics Department</p> <p>Institute of Plant Genetics and Crop Plant Research (IPK)</p> <p>Corrensstr. 3, D-06466 Gatersleben, Germany</p> <p>Tel: 0049 (0)39482-5660</p> <p>Fax: 0049 (0)39482-5595</p> <p>Email: schweizipk@gatersleben.de</p> <p>Insert Length: 682 Std Error: 0.00</p> <p>Plate: 1 row: B column: 22</p> <p>Seq primer: M13rev.</p> <p>Location/Qualifiers</p> <p>1..682</p> <p>/organism="Hordeum vulgare"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Ingrid BC mlo-5"</p> <p>/db_xref="GABI:703742"</p> <p>/db_xref="taxon:4513"</p> <p>/clone="H001B22"</p> <p>/tissue_type="leaf epidermis, 6 h and 24 h post inoculation with Blumeria graminis"</p> <p>/dev_stage="7 d after germination"</p> <p>/lab_host="XL10-Gold"</p> <p>/clone_lib="H0"</p> <p>/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Approximately 5 % of the clones correspond to cDNA from the fungi B. graminis hordei and tritici, respectively. Due to a cloning artifact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"</p>	191 a 173 c 152 g 166 t	<p>Query Match 4.5%; Score 72.2; DB 14; Length 682;</p> <p>Best Local Similarity 57.2%; Pred. No. 1.le-09;</p> <p>Matches 131; Conservative 0; Mismatches 98; Indels 0; Gaps 0;</p>
<p>187 CATACCTTCTCTGCTTCGAGCCAGACATACCAAGTCCACCTATACATG 246</p> <p> </p> <p>438 CATCTCTTTACTGGTTTTTGTAGTCCGAAACATCCGAGAACGCCAGTCATG 497</p> <p> </p> <p>247 TGGTTGAATGGTGGCCCTGGAGGAGTCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCT 306</p> <p> </p> <p>498 TGGCTCAAGGGCGCCCTGGATGTTGCTCATTAAGTGGATTATCTTGAGCTGGCCG 557</p> <p> </p> <p>307 TGGCATGCAATTCAGCTTTTGTAGCTACATACACCTCTACTCGTGGAGAGGCTTCC 366</p> <p> </p> <p>558 GCGTCAATCGATAAAGTGAAGGTGGTCAATACACCTTATCTGGAATCGAAGGCA 617</p> <p> </p> <p>367 AATTACTATCTCTGCCAGCATTTGGAGTCCGCTTTTCATATAGT 415</p> <p> </p> <p>618 TCAGTTATTTCTTGACCAACCTTTAAITGCGGATACCTGTCAGCG 666</p>	<p>187 CATACCTTCTCTGCTTCGAGCCAGACATACCAAGTCCACCTATACATG 246</p> <p> </p> <p>438 CATCTCTTTACTGGTTTTTGTAGTCCGAAACATCCGAGAACGCCAGTCATG 497</p> <p> </p> <p>247 TGGTTGAATGGTGGCCCTGGAGGAGTCTTTGATCGGTCTCTTCGAAGAGTTGGGCCCT 306</p> <p> </p> <p>498 TGGCTCAAGGGCGCCCTGGATGTTGCTCATTAAGTGGATTATCTTGAGCTGGCCG 557</p> <p> </p> <p>307 TGGCATGCAATTCAGCTTTTGTAGCTACATACACCTCTACTCGTGGAGAGGCTTCC 366</p> <p> </p> <p>558 GCGTCAATCGATAAAGTGAAGGTGGTCAATACACCTTATCTGGAATCGAAGGCA 617</p> <p> </p> <p>367 AATTACTATCTCTGCCAGCATTTGGAGTCCGCTTTTCATATAGT 415</p> <p> </p> <p>618 TCAGTTATTTCTTGACCAACCTTTAAITGCGGATACCTGTCAGCG 666</p>	<p>BASE COUNT 191 a 173 c 152 g 166 t</p> <p>ORIGIN</p>	
<p>CD054523</p> <p>CD054523.1 GI:30595544</p> <p>EST.</p> <p>KEYWORDS</p> <p>SOURCE Hordeum vulgare</p> <p>ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.</p> <p>REFERENCE 1 (bases 1 to 682)</p> <p>AUTHORS Zierold,U. and Schweizer,P.</p> <p>TITLE Barley ESTs from pathogen-attacked leaf epidermis</p> <p>COMMENT Unpublished</p> <p>Contact: Patrick Schweizer</p> <p>Transcriptome Analysis, Cytoogenetics Department</p> <p>Institute of Plant Genetics and Crop Plant Research (IPK)</p> <p>Corrensstr. 3, D-06466 Gatersleben, Germany</p> <p>Tel: 0049 (0)39482-5660</p> <p>Fax: 0049 (0)39482-5595</p> <p>Email: schweizipk@gatersleben.de</p> <p>Insert Length: 682 Std Error: 0.00</p> <p>Plate: 1 row: B column: 22</p> <p>Seq primer: M13rev.</p> <p>Location/Qualifiers</p> <p>1..682</p> <p>/organism="Hordeum vulgare"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Ingrid BC mlo-5"</p> <p>/db_xref="GABI:703742"</p> <p>/db_xref="taxon:4513"</p> <p>/clone="H001B22"</p> <p>/tissue_type="leaf epidermis, 6 h and 24 h post inoculation with Blumeria graminis"</p> <p>/dev_stage="7 d after germination"</p> <p>/lab_host="XL10-Gold"</p> <p>/clone_lib="H0"</p> <p>/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Approximately 5 % of the clones correspond to cDNA from the fungi B. graminis hordei and tritici, respectively. Due to a cloning artifact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"</p>	<p>CD054523</p> <p>CD054523.1 GI:30595544</p> <p>EST.</p> <p>KEYWORDS</p> <p>SOURCE Hordeum vulgare</p> <p>ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.</p> <p>REFERENCE 1 (bases 1 to 682)</p> <p>AUTHORS Zierold,U. and Schweizer,P.</p> <p>TITLE Barley ESTs from pathogen-attacked leaf epidermis</p> <p>COMMENT Unpublished</p> <p>Contact: Patrick Schweizer</p> <p>Transcriptome Analysis, Cytoogenetics Department</p> <p>Institute of Plant Genetics and Crop Plant Research (IPK)</p> <p>Corrensstr. 3, D-06466 Gatersleben, Germany</p> <p>Tel: 0049 (0)39482-5660</p> <p>Fax: 0049 (0)39482-5595</p> <p>Email: schweizipk@gatersleben.de</p> <p>Insert Length: 682 Std Error: 0.00</p> <p>Plate: 1 row: B column: 22</p> <p>Seq primer: M13rev.</p> <p>Location/Qualifiers</p> <p>1..682</p> <p>/organism="Hordeum vulgare"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Ingrid BC mlo-5"</p> <p>/db_xref="GABI:703742"</p> <p>/db_xref="taxon:4513"</p> <p>/clone="H001B22"</p> <p>/tissue_type="leaf epidermis, 6 h and 24 h post inoculation with Blumeria graminis"</p> <p>/dev_stage="7 d after germination"</p> <p>/lab_host="XL10-Gold"</p> <p>/clone_lib="H0"</p> <p>/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Approximately 5 % of the clones correspond to cDNA from the fungi B. graminis hordei and tritici, respectively. Due to a cloning artifact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"</p>	<p>BASE COUNT 191 a 173 c 152 g 166 t</p> <p>ORIGIN</p>	
<p>CD054523</p> <p>CD054523.1 GI:30595544</p> <p>EST.</p> <p>KEYWORDS</p> <p>SOURCE Hordeum vulgare</p> <p>ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.</p> <p>REFERENCE 1 (bases 1 to 682)</p> <p>AUTHORS Zierold,U. and Schweizer,P.</p> <p>TITLE Barley ESTs from pathogen-attacked leaf epidermis</p> <p>COMMENT Unpublished</p> <p>Contact: Patrick Schweizer</p> <p>Transcriptome Analysis, Cytoogenetics Department</p> <p>Institute of Plant Genetics and Crop Plant Research (IPK)</p> <p>Corrensstr. 3, D-06466 Gatersleben, Germany</p> <p>Tel: 0049 (0)39482-5660</p> <p>Fax: 0049 (0)39482-5595</p> <p>Email: schweizipk@gatersleben.de</p> <p>Insert Length: 682 Std Error: 0.00</p> <p>Plate: 1 row: B column: 22</p> <p>Seq primer: M13rev.</p> <p>Location/Qualifiers</p> <p>1..682</p> <p>/organism="Hordeum vulgare"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Ingrid BC mlo-5"</p> <p>/db_xref="GABI:703742"</p> <p>/db_xref="taxon:4513"</p> <p>/clone="H001B22"</p> <p>/tissue_type="leaf epidermis, 6 h and 24 h post inoculation with Blumeria graminis"</p> <p>/dev_stage="7 d after germination"</p> <p>/lab_host="XL10-Gold"</p> <p>/clone_lib="H0"</p> <p>/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Approximately 5 % of the clones correspond to cDNA from the fungi B. graminis hordei and tritici, respectively. Due to a cloning artifact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"</p>	<p>CD054523</p>		

Db	628	CAGGCTATTGAACGGCCCTTGGAGATACAGAGACTTTGACGTTTATGATATTCGGGC	569
Qy	1030	CCATATGATGACCCGACTCGCCAACTTATTACAACAATTTCTGGCAAGGACTCTGTC	1089
Db	568	CCGAGTAAGCAACCTTTTCCATCGACTTACTCGACTTACTTCAGTCAATCAAGTGC	509
Qy	1090	ATGGAGCGCTATCGGCGTCAACATCACTACACCCAGTCCAAATTAATGAGTCTACTACGT	1149
Db	508	ATGAAAGCTATTGGCGCGAGTCAAGATACGGGAATGTCCGAGGCGGTTATGACAAG	449
Qy	1150	TTCACGAAACAGGCGACTTTGTCTGGCCAACTTCATCGAAGACCTCGAGGAGATCCCT	1209
Db	448	TTTATCAACAGTGGCGATCGGGAAGCAATCTTTGTCCACATGTGCCAAG--TCATT	392
Qy	1210	GCTCTCCCGTGGTGTCCTCATCTATGGCGAGCGGATTACATCTGCAACTGGTTC	1269
Db	391	GACTCCAAGATCAAGTATGATCTGGCTGGTGGATCCGACTGGATCTGCAACTGGAT	333
Qy	1270	GGCGGTACGGCGTTCTTCCTCGTGGCGAATCTCCAGCGCCGAGTTCGGAACGCA	1329
Db	332	-----GGGTAAATACCGCGCTCTCAACTATTGCTCCCGACGTCTGCTCTCAGCT	281
Qy	1330	GGGTACAGGCCCTGAAAGTCAACGGCTGAGTATGGGAACTCCGAGTATGTAAT	1389
Db	280	CCTCTCAATCCCTTACTGTTATGGAAGTAAGTACGGAGATTTAAGACATCTGGAAT	221
Qy	1390	TTCTCTCTACTCGCTCTATGAGGCGAGGCCATGAAGTCCCATACTACCGCCCATCGGC	1449
Db	220	TTGAGCTGGCTGGAGTTATGTTGCTGGTCAIGAGTCCCSCTTATCAGCTCAGCT	161
Qy	1450	TCCCTGGAATGTTTAACGGGACTATCTCGGTGGGATATCGCA	1494
Db	160	GCATTGGCTGCTTTGTGCGCCACATTTGCCAAGAGCTATTCA	116

Search completed: September 16, 2003, 15:55:38
Job time : 4190 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 21:25:11 ; Search time 74 seconds
(without alignments)
1151.840 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555

Perfect score: 2887

Sequence: 1 LPGSTASVGRRLPKNPITG.....HTQSSVPLTATSMSSVGMCA 537

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	99.9	554	19	AAW56099
2	2338	81.0	551	23	ABR38864
3	980.5	34.0	623	23	ABR38865
4	941	32.6	554	23	ABR38859
5	916.5	31.7	526	23	ABR38817
6	636	22.0	508	23	ABR38817
7	630	21.8	536	23	ABR38819
8	630	21.8	556	17	AAR96737
9	620	21.5	557	17	AAR96738
					A. oryzae ATCC2038
					A. niger serine ca
					A. niger serine ca
					A. niger serine ca
					A. niger serine ca
					A. niger serine ca
					S. cerevisiae BAX-
					A. niger carboxype
					A. niger Bo-1 carb
					A. niger SFAG 2 ca

10	568.5	19.7	491	15	AAW48059	Sequence of protea
11	543.5	18.8	550	23	ABG93281	C. albicans BAX-as
12	414	14.3	481	23	ABR38845	A. niger carboxype
13	383	13.3	482	21	AAG30065	Arabidopsis thalia
14	383	13.3	502	21	AAG30064	Arabidopsis thalia
15	365	12.6	455	23	ABR38843	A. niger carboxype
16	361.5	12.5	479	21	AAG23905	Arabidopsis thalia
17	361.5	12.5	486	21	AAG23904	Arabidopsis thalia
18	361	12.5	476	19	AAW72966	Human serine carbo
19	361	12.5	476	22	AAW88381	Human membrae or
20	359	12.4	476	20	AAV13372	Amino acid sequenc
21	359	12.4	476	21	AAW01407	Human TANGO 176.
22	359	12.4	476	22	AAU29228	Human PRO polypt
23	359	12.4	476	22	AAW88587	Human hydrophobic
24	359	12.4	476	22	AAW80240	Novel human prote
25	359	12.4	476	24	ABU69650	Human secret
26	359	12.4	476	24	ABU71316	Human PRO223 prote
27	359	12.4	476	24	ABU71473	Human PRO polypt
28	359	12.4	476	24	ABU71919	Human secreted/tra
29	359	12.4	476	24	ABU65773	Human secreted/tra
30	359	12.4	476	24	ABU66106	Novel human secret
31	359	12.4	476	24	ABU67373	Human secreted/tra
32	359	12.4	476	24	ABU67610	Human secreted pro
33	359	12.4	476	24	ABU64527	Human secreted/tra
34	359	12.4	476	24	ABU65468	Human secreted/tra
35	359	12.4	476	24	ABU58604	Human PRO polypt
36	359	12.4	476	24	ABU56140	Human secreted/tra
37	359	12.4	476	24	ABU57135	Human PRO polypt
38	359	12.4	476	24	ABU54375	Human secreted/tra
39	359	12.4	476	24	ABU70714	Human secreted/tra
40	359	12.4	476	24	AAW25810	Human protein sequ
41	357.5	12.4	479	21	AAG47178	Arabidopsis thalia
42	357.5	12.4	486	21	AAG47177	Arabidopsis thalia
43	357	12.4	476	20	AAV28570	Secreted peptide C
44	357	12.4	476	22	AAU39043	Human secreted pro
45	357	12.4	476	23	ABW55752	Human polypeptide

ALIGNMENTS

RESULT 1

AAW56099
ID AAW56099 standard; Protein: 554 AA.

XX AAW56099;

AC (first entry)

DT 27-AUG-1998 (first entry)

XX A. oryzae ATCC20386 carboxypeptidase I protein.

XX Carboxypeptidase I; flavour improving agent; hydrolysate; proteinaceous; food industry.

XX Aspergillus oryzae.

XX Key Location/Qualifiers

XX Peptide 1..18

XX Protein /label= signal

XX /label= carboxypeptidase I

XX W09814599-AL.

XX 09-APR-1998.

XX 03-OCT-1997: 97WO-US17977.

XX 27-NOV-1996; 96US-0757534.

XX 04-OCT-1996; 96US-0726880.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO-NORDISK AS.

APP. 102b

102e

XX Berka R, Blinkovsky A, Brown K, Danbmann C, Golightly E;
PI Klotz A, Mathisen TE, Rey M;
XX WPI; 1998-240098/21.
DR N-PSDB; AAV28620.
XX Carboxypeptidase from Aspergillus oryzae - hydrolyses proteinaceous
PT substrates, useful for improving flavour of foods
XX Claim 1; Fig 3; 82pp; English.
XX This sequence represents carboxypeptidase I from Aspergillus oryzae.
CC This polypeptide has an optimal activity in the range of pH 3.0-7.5 at
CC 25 deg. C, optimal activity in the range of 55-60 deg. C at pH 4, and a
CC residual activity of at least 65.5% after 30 minutes at pH 4.0 and
CC 60 deg. C. It also has the capacity to hydrolyse X from N-Cbz-Ala-X where
CC N-Cbz is N-carboxybenzoyl and X is any amino acid. The carboxypeptidases
CC can be used for obtaining hydrolysates (which can be enriched in free
CC glutamic acid or peptide bound glutamic acid residues) from proteinaceous
CC substrates. The carboxypeptidases can be used in flavour-improving
CC compositions in the food industry. The products can also be used for the
CC production of polypeptides free of carboxypeptidase activity.
XX Sequence 554 AA;
QY Query Match 99.9%; Score 2883; DB 19; Length 554;
Best Local Similarity 100.0%; Pred. No. 1.9e-244;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 LPGSTPASVGRRLPKNPTGVKTLTANNVTIRYKEPGAEGVCETTPGVKSYSGYVDTSP 60
19 LPGSTPASVGRRLPKNPTGVKTLTANNVTIRYKEPGAEGVCETTPGVKSYSGYVDTSP 78
61 ESHTFFFFFAHNPETAPITLWNGPGSDSLGLFEELGPCHVNSTFDDYINPHSWNE 120
79 ESHTFFFFFAHNPETAPITLWNGPGSDSLGLFEELGPCHVNSTFDDYINPHSWNE 138
121 VSNLLFLSQPLGVGFSYSDVVGSIINPTGVVNSSFAGVQGYPTDITLDTNLAEE 180
139 VSNLLFLSQPLGVGFSYSDVVGSIINPTGVVNSSFAGVQGYPTDITLDTNLAEE 198
181 AAWETLQGLSLPSLRVQSKDFSLWTSYGGHYGPAFFNHFYEQNERIANGSVANGVQ 240
199 AAWETLQGLSLPSLRVQSKDFSLWTSYGGHYGPAFFNHFYEQNERIANGSVANGVQ 258
241 LNFNSLGIINGIIDEAIOAPYYPFAVNNIYGIKAVNETVYNNYKMFANQMPNGCQDLIST 300
259 LNFNSLGIINGIIDEAIOAPYYPFAVNNIYGIKAVNETVYNNYKMFANQMPNGCQDLIST 318
301 CKQTNRTALADYALCAEATNMCNDRNVEGPPYAFAGRYDIDRHPYDDPTPPSYNKKFLAK 360
319 CKQTNRTALADYALCAEATNMCNDRNVEGPPYAFAGRYDIDRHPYDDPTPPSYNKKFLAK 378
361 DSVMDAIGVNIYQSNNDVYAFQQTGDFWPNFIEDLEILALPVKVSILYGDADYIC 420
379 DSVMDAIGVNIYQSNNDVYAFQQTGDFWPNFIEDLEILALPVKVSILYGDADYIC 438
421 NWFQGGQVSLAANTYSQAQFRSAGYTPKLVNGVEYGETREYCNFSFRVYAGHEVPVYQ 480
439 NWFQGGQVSLAANTYSQAQFRSAGYTPKLVNGVEYGETREYCNFSFRVYAGHEVPVYQ 498
481 PIASLQLPNRTIFGWDIAEGOKKIWPSTKTTGTATATHTQSSVPLPTATSMSSVGM 536
499 PIASLQLPNRTIFGWDIAEGOKKIWPSTKTTGTATATHTQSSVPLPTATSMSSVGM 554
RESULT 2
ABR38864
ID ABR38864 standard; Protein; 551 AA.
XX
AC ABR38864;
XX

DT XX
DE XX
XX XX
KW KW
KW EC3.4.16.6.
XX ASpergillus niger.
XX OS
XX WO200268623-A2.
XX
XX 06-SEP-2002.
XX 22-FEB-2002; 2002WO-EP01984.
XX 23-FEB-2001; 2001EP-0200657.
XX 23-FEB-2001; 2001EP-0200658.
XX 23-FEB-2001; 2001EP-0200660.
XX 26-FEB-2001; 2001EP-0200706.
XX 26-FEB-2001; 2001EP-0200707.
XX 26-FEB-2001; 2001EP-0200708.
XX 26-FEB-2001; 2001EP-0200719.
XX 28-MAR-2001; 2001EP-0000075.
XX 28-MAR-2001; 2001EP-0000078.
XX 28-MAR-2001; 2001EP-0000080.
XX 28-MAR-2001; 2001EP-0000087.
XX 28-MAR-2001; 2001EP-0000088.
XX 21-MAY-2001; 2001EP-0000156.
XX 21-MAY-2001; 2001EP-0000159.
XX 21-MAY-2001; 2001EP-0000160.
XX 21-MAY-2001; 2001EP-0000162.
XX 21-MAY-2001; 2001EP-0000165.
XX 21-MAY-2001; 2001EP-0000166.
XX 21-MAY-2001; 2001EP-0000168.
XX 21-JUN-2001; 2001EP-0000240.
XX 21-JUN-2001; 2001EP-0000242.
XX 21-JUN-2001; 2001EP-0000244.
XX 21-JUN-2001; 2001EP-0000246.
XX 12-JUL-2001; 2001EP-0000280.
XX 12-JUL-2001; 2001EP-0000285.
XX 30-JUL-2001; 2001EP-0000323.
XX 30-JUL-2001; 2001EP-0000327.
XX 02-AUG-2001; 2001EP-0000341.
XX 02-AUG-2001; 2001EP-0000342.
XX 02-AUG-2001; 2001EP-0000343.
XX 02-AUG-2001; 2001EP-0000344.
XX 09-AUG-2001; 2001EP-0000357.
XX 16-AUG-2001; 2001EP-0000374.
XX 16-AUG-2001; 2001EP-0000377.
XX 20-SEP-2001; 2001EP-0000478.
XX 20-SEP-2001; 2001EP-0000483.
XX 22-OCT-2001; 2001EP-0000552.
XX 22-OCT-2001; 2001EP-0000553.
XX 22-OCT-2001; 2001EP-0000554.
XX 22-OCT-2001; 2001EP-0000556.
XX 22-OCT-2001; 2001EP-0000557.
XX 22-OCT-2001; 2001EP-0000558.
XX 15-NOV-2001; 2001EP-00004464.
XX 21-DEC-2001; 2001EP-0005117.
XX (STAM) DSM NV.
XX Edens L, Van Dijk AA, Krubasik P, Alberman K, Stock A, Kimpel E;
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
PI Stiebler J, Albarg R;
XX WPI; 2002-723203/78.
DR P-PSDB; ABZ78231, ABZ78288.
XX
XX Novel isolated protease polypeptide useful in laboratory, clinical,
PI pharmaceutical, chemical, diagnostic, personal care and industrial

24-APR-2003 (first entry)
A. niger serine carboxypeptidase polypeptide #3.
Protease; fungal infection; aspergillosis; food; tanning; detergent;
protein solubility; viscosity; taste; texture; nutritional value;
EC3.4.16.6.
Aspergillus niger.
WO200268623-A2.
06-SEP-2002.
22-FEB-2002; 2002WO-EP01984.
23-FEB-2001; 2001EP-0200657.
23-FEB-2001; 2001EP-0200658.
23-FEB-2001; 2001EP-0200660.
26-FEB-2001; 2001EP-0200706.
26-FEB-2001; 2001EP-0200707.
26-FEB-2001; 2001EP-0200708.
26-FEB-2001; 2001EP-0200719.
28-MAR-2001; 2001EP-0000075.
28-MAR-2001; 2001EP-0000078.
28-MAR-2001; 2001EP-0000080.
28-MAR-2001; 2001EP-0000087.
28-MAR-2001; 2001EP-0000088.
21-MAY-2001; 2001EP-0000156.
21-MAY-2001; 2001EP-0000159.
21-MAY-2001; 2001EP-0000160.
21-MAY-2001; 2001EP-0000162.
21-MAY-2001; 2001EP-0000165.
21-MAY-2001; 2001EP-0000166.
21-MAY-2001; 2001EP-0000168.
21-JUN-2001; 2001EP-0000240.
21-JUN-2001; 2001EP-0000242.
21-JUN-2001; 2001EP-0000244.
21-JUN-2001; 2001EP-0000246.
12-JUL-2001; 2001EP-0000280.
12-JUL-2001; 2001EP-0000285.
30-JUL-2001; 2001EP-0000323.
30-JUL-2001; 2001EP-0000327.
02-AUG-2001; 2001EP-0000341.
02-AUG-2001; 2001EP-0000342.
02-AUG-2001; 2001EP-0000343.
02-AUG-2001; 2001EP-0000344.
09-AUG-2001; 2001EP-0000357.
16-AUG-2001; 2001EP-0000374.
16-AUG-2001; 2001EP-0000377.
20-SEP-2001; 2001EP-0000478.
20-SEP-2001; 2001EP-0000483.
22-OCT-2001; 2001EP-0000552.
22-OCT-2001; 2001EP-0000553.
22-OCT-2001; 2001EP-0000554.
22-OCT-2001; 2001EP-0000556.
22-OCT-2001; 2001EP-0000557.
22-OCT-2001; 2001EP-0000558.
15-NOV-2001; 2001EP-00004464.
21-DEC-2001; 2001EP-0005117.

CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,
CC cheese manufacturing, meat tenderizing), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABR38413-ABR38869 represent the A. niger proteases of
CC the invention.

XX SQ Sequence 623 AA;

Query Match 34.0%; Score 980.5; DB 23; Length 623;
Best Local Similarity 39.8%; Pred. No. 4e-77;
Matches 216; Conservative 91; Mismatches 178; Indels 61; Gaps 19;

QY 26 TANNVTIRYKE--PGAE-----GVCETTPGVKSYGVDTSPES----- 62
DB 32 TPEDLVIISEIFPGARISYKQPLGICTTPTSPSYSGYIHLPHWILTNLSIRGISISQ 91
QY 63 ---HTFFWFPEAR--HNPTAPITELMNGPGSDSLIGLFEELGCPCHVNSTFDD--YINPH 116
DB 92 YPINTFFWYPPSRHHNNHNDTSPLTIWNGGPGGSSMIGLFGQENGPCVTNEDSNSTAYNPW 151
QY 117 SNNEVSNLLFSLQPLGVGSYSTVDGSIINPTGVVYVENSFAGVQRYPIID--ALIDIT 174
DB 152 SNWEYVDMLYIEQPVQTGFSGYDVLNRGTLD-----LNETF--LVGTLPSQDVHGVTNGT 203
QY 175 TNLAAAEAAWELLQGLFSG----LPSLDSRVQSKD-FSLWTESYGGHYPAFENHFYBONE 229
DB 204 VN-CGRALWVALQVWIGFSEYSSVDNGCGGDDRVSIWTESYGGYGRAYTALFQEMNE 262
QY 230 RIANGSVN-GVQLNFSNLGIINGIDETQAPYYPEFAVNNTYGIKAVNETVYNYMKFAN 288
DB 263 RIESGEVSTGKTHLDTLGLINGCVLLVQVPSFEQAYNNTYGIENINTLYDRAMDSW 322
QY 289 QMENGCDLSTCKQINRTALAD-YALCAEATNMCQDNVEGPPYAFAGRGVYDIRHPYD 347
DB 323 SKPGGGRDMFIECKRAGE--LGQPLICEASDYCSREIKSLYNTISGRGYDIAHPTPD 380
QY 348 PTPPSYNKFELAKSDVDAINNTKQSNNDVYAFQQTGFQWVP-----NFIEDLEEI 402
DB 381 AALVPYFVGFNLRPWQKALGVVNYTMSSEAVGNSFASTGDY--PRNDPRGMIGDIGYL 438
QY 403 LALPVRVSLIYGADYICNWFQGVQAVSLAANYSCAAQFRSAGYTPDLKVNQVEY--GETRE 460
DB 439 LDSGVKVMYIGRDYACPRWGGEVSLIWEYDAEKFRAGYAEVQTKS--SYVGGILVRQ 497
QY 461 YGNFSTRVYAGHEVPTYOPIASLOLFNRTIFGWDIAEGQKKIW--PSYKTINGTAT 518
DB 498 YGNFSTRVYQAGHEVFPYQETAYEILFNRAQENWMDIAGGISLEQNSYGTGSPSTW 557
QY 519 TQSSVP 524
DB 558 IKNEVP 563

RESULT 4
ABR38859
ID ABR38859 standard; Protein: 554 AA.

XX AC ABR38859;

XX DT 24-APR-2003 (first entry)

XX A. niger serine carboxypeptidase polypeptide #2.

XX Protease; fungal infection; aspergillosis; food; tanning; detergent;
KW protein solubility; viscosity; taste; texture; nutritional value;

KW EC3.4.16.6.
XX Aspergillus niger.
XX WO200268623-A2.
XX 06-SEP-2002.
XX 22-FEB-2002; 2002WO-EP01984.
XX 23-FEB-2001; 2001EP-0200657.
XX 23-FEB-2001; 2001EP-0200658.
XX 23-FEB-2001; 2001EP-0200660.
XX 26-FEB-2001; 2001EP-0200706.
XX 26-FEB-2001; 2001EP-0200707.
XX 26-FEB-2001; 2001EP-0200708.
XX 26-FEB-2001; 2001EP-0200719.
XX 28-MAR-2001; 2001EP-0000075.
XX 28-MAR-2001; 2001EP-0000078.
XX 28-MAR-2001; 2001EP-0000080.
XX 28-MAR-2001; 2001EP-0000087.
XX 28-MAR-2001; 2001EP-0000088.
XX 21-MAY-2001; 2001EP-0000156.
XX 21-MAY-2001; 2001EP-0000159.
XX 21-MAY-2001; 2001EP-0000160.
XX 21-MAY-2001; 2001EP-0000162.
XX 21-MAY-2001; 2001EP-0000165.
XX 21-MAY-2001; 2001EP-0000166.
XX 21-MAY-2001; 2001EP-0000168.
XX 21-JUN-2001; 2001EP-0000240.
XX 21-JUN-2001; 2001EP-0000242.
XX 21-JUN-2001; 2001EP-0000244.
XX 12-JUL-2001; 2001EP-0000246.
XX 12-JUL-2001; 2001EP-0000280.
XX 12-JUL-2001; 2001EP-0000285.
XX 30-JUL-2001; 2001EP-0000323.
XX 30-JUL-2001; 2001EP-0000327.
XX 02-AUG-2001; 2001EP-0000341.
XX 02-AUG-2001; 2001EP-0000342.
XX 02-AUG-2001; 2001EP-0000343.
XX 02-AUG-2001; 2001EP-0000344.
XX 09-AUG-2001; 2001EP-0000357.
XX 16-AUG-2001; 2001EP-0000374.
XX 16-AUG-2001; 2001EP-0000377.
XX 20-SEP-2001; 2001EP-0000478.
XX 20-SEP-2001; 2001EP-0000483.
XX 22-OCT-2001; 2001EP-0000552.
XX 22-OCT-2001; 2001EP-0000553.
XX 22-OCT-2001; 2001EP-0000554.
XX 22-OCT-2001; 2001EP-0000556.
XX 22-OCT-2001; 2001EP-0000557.
XX 22-OCT-2001; 2001EP-0000558.
XX 15-NOV-2001; 2001EP-0000464.
XX 21-DEC-2001; 2001EP-00005117.
XX (STAM) DSM NV.

Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel E;
Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
Maier D, Spreafico F, Folkers U, Hopper S, Kemmer W, Tan P;
Stiebler J, Albarg R;

WFI; 2002-723203/78.
P-PSDB; ABZ78226, ABZ78283.

Novel isolated protease polypeptide useful in laboratory, clinical,
pharmaceutical, chemical, diagnostic, personal care and industrial
applications -

Claim 13; Page 366-369; 394pp; English.

The invention relates to a novel isolated protease polypeptide. A
polypeptide or polynucleotide of the invention is useful for diagnosing a

CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing, in
CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABR38313-ABR38869 represent the A. niger proteases of
CC the invention.

CC a search against public databases. A polypeptide of the invention is
 CC useful in a selected number of industrial or pharmaceutical processes, in
 CC laboratory or clinical processes, in food industry (baking, brewing,
 CC cheese manufacturing, meat tenderising), in tanning industry and in the
 CC manufacture of biological detergents. A polypeptide may also be useful
 CC for improving protein solubility, extraction yields, viscosity or taste,
 CC texture, nutritional value, minimising of antigenicity or
 CC anti-nutritional factors, colour or functionality as well as processing
 CC aspects like filterability of the proteinaceous raw material. The
 CC sequences shown in ABR38313-ABR38869 represent the *A. niger* proteases of
 CC the invention.

XX SQ Sequence 526 AA;

Query Match 31.7%; Score 916.5; DB 23; Length 526;
 Best Local Similarity 38.0%; Pred. No. 1.3e-71;
 Matches 212; Conservative 78; Mismatches 187; Indels 81; Gaps 13;

QY 9 VGRQLPKNPTGVKTLTANNVTIRYKEPAGVCEITPGVKSYYVDTPESHTFFWF 68
 DB 16 VSRAQFVAPPDPLITKGYLDIPVKYQV-PTGICETDPSVKSFSGYVDVAEHSHTFFWF 74
 QY 69 FEARH-NPETAPIILWNG-----GPGSDSLIGLFEELGCPCHVNSTFDYINPHSNVEVS 122
 DB 75 FEARNQDTEAPLTVWINGMSDFGPGSSMIGLFQEHGCGCIDANGSVYNNPYSWNNAS 134
 QY 123 NLLFLSOPLGVSFSDTVDSGINPVTGVVENSFGVQGRYPITDAFLDITNLAAEAA 182
 DB 135 NMLYDQVQVGFYS-----IPVGYVDSSTNGEMGAPP----- 170
 QY 183 WEILQGLSLGLSDRSVQSKDFSLWTESYGGHYGPAFFNHFYEONERIANGSVNGVOLN 242
 DB 171 -----QYSRETFHTTESYGGHYGVPNEVTEQNAHLQPCA---KKIQ 211
 QY 243 FNSLGIINGLIIDEAQIAPYEPFAV--NNTYGIKAVNETVYNYMKFANQMENGCDLIST 300
 DB 212 LGSVMIGNGYDPIIQYQAYNFTVPGNTYDYLDFNKSISLSLNNLYGPGNCJLDYD 271
 QY 301 CKQNTALADYALCAEATNMCRONVGPYAFAGRGYDIDRHPYDDPTPSSYNNKFLAK 360
 DB 272 C-----ARGIDEICSTADRCANEVENVYDIYSGRDEYDFRELTDFPFYFYDYLNK 326
 QY 361 DSVMDAIGVNNYTONSDNYVAFQOTG-DFWPNFIEDLEILALPVKVSILYGDAYI 419
 DB 327 ASVQAAGIAYINYESNNAVLAFSGTGDGRIMNTIQDVGKLLKQGVTVVMYAGDADYN 386
 QY 420 CNWFGSOAVSLAANYSOAAQPSAGYPLKV-NGVEYGETREYGFSTRVYEAGHEVPY 478
 DB 387 CNWLGEAVSLQV---KAANFSSAGYTNIVTSDGVTGQVRQAGQFAFVRVYESGHEVPF 443
 QY 479 YQPIASLQLENETIFGWDIAEQKKIWPSSYKTNGT-----ATATH-T 519
 DB 444 YQPLALEMFERVIGKQVATCKIPISSSLQVGTGPKSYRBSNSTIQWEVLDSLATYNT 503
 QY 520 OSSVPLPATSSSSVGM 537
 DB 504 TTNAPNPVSRRLKRMGA 521

RESULT 6

ABG93049
 ID ABG93049 standard; Protein; 508 AA.

XX AC ABG93049;

XX DT 21-NOV-2002 (first entry)

XX DE S. cerevisiae BAX-associated protein fragment SEQ ID 56.

XX KW Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
 KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
 KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
 KW neurodegeneration; cell death.

XX OS Saccharomyces cerevisiae.
 XX PN WO200264766-A2.
 XX PD 22-AUG-2002.
 XX PF 21-DEC-2001; 2001WO-EP15398.
 XX PR 22-DEC-2000; 2000EP-0870318.
 XX PR 04-JAN-2001; 2001EP-0870002.
 XX PR 09-JAN-2001; 2001EP-0870003.
 XX PA (JANC) JANSSEN PHARM NV.
 XX PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
 XX WPI; 2002-667002/71.
 XX DR N-PSDB; ABQ76315.
 XX CC New isolated nucleic acid representing a synthetic Bax gene, useful as
 CC medicament for treating, preventing and/or alleviating yeast or fungal
 CC infections or proliferative disorders, or for preventing apoptosis in
 CC certain diseases -
 XX PS Claim 36; Figure 1; 344pp; English.
 XX CC This invention describes a novel nucleic acid representing a synthetic
 CC Bax gene. The Bax gene of the invention is useful for identifying
 CC Bax-resistant yeast or fungi, identifying, or obtaining and identifying
 CC Candida spp. sequences that are differentially expressed in a pathway
 CC eventually leading to programmed cell death or identifying inhibitors or
 CC inhibitor sequences of Bax-induced cell death. The products of the
 CC invention have cytostatic, fungicide; immunosuppressive, virucide and
 CC vasotropic activity and can be used in vaccines or for gene therapy. The
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
 CC antisense molecules and antibodies are useful as medicaments or in
 CC preparing a medicament for treating, preventing and/or alleviating
 CC diseases associated with yeast or fungi or proliferative disorders, such
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds
 CC or polypeptides, or the genetically modified organism are useful for
 CC preparing a medicament for modifying the endogenic flora of humans and
 CC other mammals. The vaccine is useful for immunising against yeast or
 CC fungal infections. Apoptosis-related diseases include autoimmune disease,
 CC ischaemia, diseases related with viral infections or neurodegenerations.
 CC This sequence represents a polypeptide associated with the Bax gene
 CC described in the disclosure of the invention.

XX SQ Sequence 508 AA;

Query Match 22.0%; Score 636; DB 23; Length 508;
 Best Local Similarity 32.6%; Pred. No. 5.6e-47;
 Matches 167; Conservative 70; Mismatches 197; Indels 78; Gaps 14;

QY 1 LPGSTPASVGRRLQ-----PKNPTGVKTLTANNVTIRYKEPAGVCEITPGVKSYSGV 56
 DB 43 LPQNTQOTLKLDRLNHDDPLFTFTTSSVDITSLRLRTVDPSKLG1----DTVKQSGYM 98
 QY 57 DTSPESTHFFWFFEARHNPETAPITLWNGPGSDSLIGLFEELGCPCHVNSTFDYINPH 116
 DB 99 DYKDSKHFFYWFESNDPANDPILWNGGPCSSFTGLILFELGSSIGADMKPITHNY 158
 QY 117 SNNEVSNLLFTSQPLGVGSYSYDVGSIINPVTGVVENSFGVQGRYPITDAFLDITFN 176
 DB 159 SWNNNASMTIFLEQLPLGVGSYGD-----EKYSSTK 188
 QY 177 LAEAAWELIQLGLSLGLSDRSVQSKDFSLWTESYGGHYGPAFFNHFYEON-ERIANGS 235
 DB 189 LACKDAYIFLELFEAFPHL-----RSDNFHAGESYAGHYIPIAHEIVVKNFER----- 239
 QY 236 VNGVQLNFNSLGIINGIIDEAQIAPYEPFAV--NNTYGIKAVNETVYNYMKFANQMENG 294
 DB 240 ----TFLNLSVMIGNGITDPLQADYEPMACGKGVPVLSSECEKCKSKAAGR----- 290

New Kluyveromyces yeast with modified protease gene - esp. used for high yield prodn. of recombinant protein, also DNA encoding yeast protease and derived peptide(s)

Disclosure; Page 28-31; 49pp; English.

The protease gene is to be modified in order to render it (partially) incapable of producing the natural protein; or result in a non-functional protease or in a protease with modified proteolytic activity. The modifications can be introduced in vitro or in situ by standard genetic engineering techniques or by exposure to mutagenic agents.

CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 491 AA;

Query Match 19.7%; Score 568.5; DB 15; Length 491;
Best Local Similarity 31.18; Pred. No. 4.6e-41;
Matches 151; Conservative 66; Mismatches 184; Indels 85; Gaps 14

QY 27 ANNVTIRKPGAGVCETTPGVKSGSYGYDTSPESTHFFWFPEARHNPETAPITLWLN 86
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 132 AYDLRVAKTDPSSIGI---DPGKQKTGYLDNDENKHLYFWFFESRNPENDPVVLWLN 188
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 86 GGPCSDSLGLFEELGPCCHVNSTFDDYINPHSNVEYNLLFLSOPLGCVGSYDVDSGI 145
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 189 GGPCSLSLTGLFELGPSSINKKIOPYNIDYAWNSASVIFLDQPNVGISYSNSA--- 244
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 146 NPTGVVNSSFAGVQGGRYPDTIDATLIDTNLAEEAAEWELLQGLFSGLDSRVCSKDF 205
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 245 -----VSOTVAAGKDVKYALLTLFPKQFP-----EYAKQDF 274
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 206 SLWTESYGHHGPAFFNHFEQNERTANGSVGNLFNSLGTINGIIDEAIOAPYEPF 265
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 275 HIAGESYAGHYIPVFASEILLSHKR-----NINLOSVLINGLTDGLTEYEYRPM 325
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 266 AVNTTYGIKAVNETVNYMKFANOMPNGCODLISTCKQINRTALADYALCAEATNMCRDN 325
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 326 ACGDG-GYPAVLDE-GSCQAMDNALPR-COSMTESCYSESA-----WVCVPASYCNNA 377
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGNIN-YTQSND 379
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 378 LLAPYQR-TGQNVYDRGKCEDSSNLCSAMGVSDYLNKTEIVIAVGAEVNGYDCNF 436
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 380 VYFAFOQTGFVWP--NFIEDLEILALPVRSLIY-GDAFYICNWFGGQAVSLAANYSQ 436
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 437 INRNFLHGDMKPYYRLVPLGLE--QIEPV--LIYAGDADFICNWLGNKAWELEWPG 491
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 437 AAQFRSAGYTPLKV-----NGVEYGETREYGNFSFTRYVEAGHEVPPYOPTIASLOLFNRT 491
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 492 QAEVASAKLEDLVVENEHKKIGQVKSNGHFTMRLYGGGHVMYMPDQPSSLEFNWR 551
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
QY 492 IFG-W 495
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
DB 552 LGGEW 556
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :

RESULT 10
AAR48059
ID AAR48059 standard; Protein; 491 AA.
AC AC
XZ XZ
XX XX
DT DT
DT DT
XX XX
DE DE
KW KW
KW KW
OS OS
XX XX
PN PN
XX XX
PD PD
PF PF
XX XX
PR PR
XX XX
PA PA
PI PI
XX XX
DR DR
DR DR
XX XX

Sequence of protease C encoded by gene K1.PRC1.
Protease; Yeast; proteolysis.
Kluyveromyces lactis.
WO9400579-A1.
06-JAN-1994.
23-JUN-1993; 93WO-FR00623.
25-JUN-1992; 92FR-0007785.
(RHON) RHONE POULENC RORER SA.
Fleer R, Fournier A, Yeh P;
MPI; 1994-026215/03.
N-PSDB; AA055347.
Candida albicans.
Bax: Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
KW apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW neurodegeneration; cell death.
C. albicans BAX-associated protein fragment SEQ ID 520.
ABG93281 standard; Protein; 550 AA.
ABG93281;
21-NOV-2002 (first entry)
C. albicans BAX-associated protein fragment SEQ ID 520.

XX WO200264766-A2.
XX 22-AUG-2002.
XX 21-DEC-2001; 2001WO-EP15398.
XX 22-DEC-2000; 2000EP-0870318.
XX 04-JAN-2001; 2001EP-0870002.
XX 09-JAN-2001; 2001EP-0870003.
XX (JANC) JANSSEN PHARM NV.
XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX WPI; 2002-667002/71.
XX N-ESDB; ABQ76547.
XX New isolated nucleic acid representing a synthetic Bax gene, useful as
XX medicament for treating, preventing and/or alleviating yeast or fungal
XX infections or proliferative disorders, or for preventing apoptosis in
XX certain diseases
XX Claim 36; Figure 2; 344pp; English.
XX This invention describes a novel nucleic acid representing a synthetic
XX Bax gene. The Bax gene of the invention is useful for identifying
XX Bax-resistant yeast or fungi, identifying, or obtaining and identifying
XX Candida spp. sequences that are differentially expressed in a pathway
XX eventually leading to programmed cell death or identifying inhibitors or
XX inhibitor sequences of Bax-induced cell death. The products of the
XX invention have cytostatic, fungicide, immunosuppressive, virucide and
XX vasotropic activity and can be used in vaccines or for gene therapy. The
XX isolated nucleic acids, polypeptides, pharmaceutical compositions,
XX antisense molecules and antibodies are useful as medicaments or in
XX preparing a medicament for treating, preventing and/or alleviating
XX diseases associated with yeast or fungi or proliferative disorders, such
XX as cancer, or for preventing apoptosis in certain diseases. The compounds
XX or polypeptides, or the genetically modified organism are useful for
XX preparing a medicament for modifying the endogenic flora of humans and
XX other mammals. The vaccine is useful for immunising against yeast or
XX fungal infections. Apoptosis-related diseases include autoimmune disease,
XX ischaemia, diseases related with viral infections or neurodegenerations.
XX This sequence represents a polypeptide associated with the Bax gene
XX described in the disclosure of the invention.
XX Sequence 550 AA;
Query Match 18.8%; Score 541.5; DB 23; Length 550;
Best Local Similarity 31.3%; Pred. No. 1.3e-38;
Matches 150; Conservative 63; Mismatches 157; Indels 109; Gaps 17;
QY 49 VKSYGYVD-TSPESHTEFWFEARHNHETAPITLWNGGPGSDSLIGLFEELGPGCHVNS 107
DB 133 VKQYTGILDIIDSLDKHLEYWFTEFNRDKNPDIILWNGGPGCSSTGLFELGPGSSINK 192
QY 108 TEDDYINSHSVENSNLLFLOPLGVGFSYSTDVDSINPVTVGVVENSFAGVQGRYPTI 167
DB 193 TLHPYINPISWNSASVIFLDQPVGVGYSY-----TCGDE----- 227
QY 168 DATLIDTTNLAABAAWEILQGLSLPSLDKSLVQSKDFSLWTSYSGHYGPAFFNFYEQ 227
DB 228 -----VKNTLTAAKDVVLELFQKFP-----QFLTNKPHIAGESYAGHYIPAF----- 272
QY 228 NERLANGSVQVLNENSLGII--NCIIDEATQAPYYPEFVAVNTYGIKAVNETVYNYMK 285
DB 273 -----ASEIINNADRSFEIASVLINGITDPLIDQGSY----- 305
QY 286 FANQMPNGC-----QDLIST--CKQTNRTALADYALCATNMCRD-----NVEGFY-- 330
DB 306 ----KPMGCGEGYKPEVLTEQDQMER----DYPRCAKLTKLCYFSQALICVPAQYIC 357
QY 331 -----YAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGV-NIN-YTOS 376

Db 358 DSRLFQPYAQGLNPYDIRKDCAGGNGCYVEMDYLDLEYLMLDYVKEAVGASNDIDFTSC 417
QY 377 NNDVYAFQQTGDFWENFIEDLEILALPVRVSLIYGDADYICNWFEGGQAVSLAANYSQ 436
Db 418 DDTVFRNFILGDEMKP-FQYVVAELLDNNVPVLLIYAGDKDYICNWLGNLAWVNELEYSD 476
QY 437 AAQFRSAGYTPKLVNGVE-YGETREYGNFSTRVYVYAGHEVYPTASLOLFNRTIFG 494
Db 477 SEHFAPKPLQWKQDGKKAAGEVKNHKTFLRIYDAGHMVFPDQPENALSMVNTWVQGS 535

RESULT 12
ABR38845
ID ABR38845 standard; Protein: 481 AA.
XX ABR38845;
AC ABR38845;
XX 24-APR-2003 (first entry)
DT A. niger carboxypeptidase Y polypeptide #3.
DE Protease; fungal infection; aspergillosis; food; tanning; detergent;
XX protein solubility; viscosity; taste; texture; nutritional value;
KW EC3.4.16.5.
XX Aspergillus niger.
XX WO200268623-A2.
PD 06-SEP-2002.
XX 22-FEB-2002; 2002WO-EP01984.
XX 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 23-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.

PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139464.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139465.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151080.
PR	18-JUN-1999;	99US-0139763.	PR	30-AUG-1999;	99US-0151303.
PR	21-JUN-1999;	99US-0139817.	PR	31-AUG-1999;	99US-0151438.
PR	22-JUN-1999;	99US-0139899.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140353.	PR	07-SEP-1999;	99US-0152363.
PR	23-JUN-1999;	99US-0140354.	PR	10-SEP-1999;	99US-0153070.
PR	24-JUN-1999;	99US-0140354.	PR	13-SEP-1999;	99US-0153758.
PR	24-JUN-1999;	99US-0140695.	PR	15-SEP-1999;	99US-0154018.
PR	28-JUN-1999;	99US-0140823.	PR	16-SEP-1999;	99US-0154039.
PR	29-JUN-1999;	99US-0140991.	PR	20-SEP-1999;	99US-0154779.
PR	30-JUN-1999;	99US-0141287.	PR	22-SEP-1999;	99US-0155139.
PR	01-JUL-1999;	99US-0141842.	PR	23-SEP-1999;	99US-0155486.
PR	01-JUL-1999;	99US-0142154.	PR	23-SEP-1999;	99US-0155659.
PR	02-JUL-1999;	99US-0142055.	PR	24-SEP-1999;	99US-0156458.
PR	06-JUL-1999;	99US-0142390.	PR	28-SEP-1999;	99US-0156596.
PR	08-JUL-1999;	99US-0142803.	PR	29-SEP-1999;	99US-0157117.
PR	09-JUL-1999;	99US-0142920.	PR	04-OCT-1999;	99US-0157755.
PR	12-JUL-1999;	99US-0143542.	PR	05-OCT-1999;	99US-0157865.
PR	13-JUL-1999;	99US-0143624.	PR	06-OCT-1999;	99US-0158029.
PR	14-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0158232.
PR	15-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158369.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158369.
PR	16-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	26-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	26-OCT-1999;	99US-0161361.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161922.
PR	04-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147192.	PR	29-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147260.	PR	29-OCT-1999;	99US-0162142.
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			

Query Match

13.3%; Score 383; DB 21; Length 482;

Best Local Similarity 26.2%; Pred. No. 9e-25;

Matches 135; Conservative 68; Mismatches 202; Indels 110; Gaps 18;

QY

15 PKNPTGKVLTTANNVILRYKE-----PGAGVCEITPG-----VKSYSGYDVTSPESHT 64

Db

32 PSPKRGVSSGDTSHENVIORESVFPKDKDLIQQLPGQFSDVTFKQYGGYVAVNPKAGR 91

QY

65 F--FWFFEARHNPETAPITLWLNKGPGSDSLIGLFEELGPGCHVNSTFDO-YINPHSWNEV 121

RESULT 1.4

PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147038.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147204.			
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	27-AUG-1999;	99US-0151303.			
PR	30-AUG-1999;	99US-0151438.			
PR	31-AUG-1999;	99US-0151930.			
PR	01-SEP-1999;	99US-0152363.			
PR	07-SEP-1999;	99US-0153070.			
PR	10-SEP-1999;	99US-0153758.			
PR	13-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	28-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	08-OCT-1999;	99US-0158232.			
PR	12-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	14-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			
PR	18-OCT-1999;	99US-0159584.			
PR	21-OCT-1999;	99US-0160741.			
PR	21-OCT-1999;	99US-0160767.			
PR	21-OCT-1999;	99US-0160768.			
PR	21-OCT-1999;	99US-0160770.			

Query Match 13.3%; Score 383; DB 21; Length 502;

Best Local Similarity 26.2%; Pred. No. 9.5e-25;

Matches 135; Conservative 68; Mismatches 202; Indels 110; Gaps 18;

QY	15	PKNPTGVKLTITANNVTIRKE---PGAGVGEITPG-----VKSYSYGVDTSPESH 64
Db	52	PSPKRGVSSGSDTSHENVIQRESVPSFKOKDLIQLPQSDVTFKQYGGYAVNKPAGR 111
QY	65	F--FWFPEAHNPETAPITLWNGPGSDSLIGLFELGCPCHVNSTFDD--YINPHSWNEV 121
Db	112	FLYFYFVETIKPGNTTPLVIFNFGPGCSSLGAFKELGFPFRVHSDGKILFRNPYSWNE 171
QY	122	SNLLFLSQPLGVGFSYSDTVDGSINFVTGVVENSFAGVGQRYPTIDAILIDTTLAAEA 181
Db	172	ANVLETPVGTGFSYS-----NSPINGKQDKAT-----AED 204
QY	182	AWELQGLSLGFLSDRSVQSKDFSLWTESYGGHYGPAFTNHFYEQNERIANGSVNGVQL 241
Db	205	NMFVLNWLRFEPPEYKGR---DIYIAGQSYAGHYVPQLAQIILHRN-----NOTLI 252
QY	242	NFNSLGIINGIIDEAIOAPYYPFAVNTYGIKAVNETVYNYKFAQMPNGCODLSTC 301
Db	253	NLRGILLGNFSLNREIQDDFGYKFMF--SHGLIS-QQOMDNYNKF-----CTD--SDL 300
QY	302	KQTNRTALADYALCAEATNMCRONVEGPIYAFAGRGVYDIRHPY----- 345
Db	301	YDWDKCHLASQKTEAQKTHL-----DIYIYAPCLNLSLSEPKKCTTI 345
QY	346	--DDPTPPSYNKFIAKDSVMDALGVN-----INTOSNNDVYAFQGTGDFVWPNI 398
Db	346	MKADPCSGNYLKAYLNIKEVQEAHANTTKIPYEWTSCTNKLWENKEDRYV--SLTPI 403
QY	399	LEETLALPVRVSLIYGADYICNWFQGOAVSLAANYSCAAQFRSAGYPLKVNVEYGET 458
Db	404	LQELMGKVRMLYNGDVLVIFPTSTLAVVYKTMNLIVVKEWR-----PWFETGSHVGFT 458
QY	459	REY-GNFSTRVYEAGHEVYQPIASLQLFNRTI 492
Db	459	EDYKGNLTFVTVKGAGHSVPTDQPIHALNIFTSEI 493

RESULT 15

ABR38843

ID ABR38843 standard; Protein; 455 AA.

XX

AC ABR38843;

XX

DT 24-APR-2003 (first entry)

XX

DE A. niger carboxypeptidase Y polypeptide #2.

XX

KW Protease; fungal infection; aspergillois; food; tanning; detergent;

KW protein solubility; viscosity; taste; texture; nutritional value;

KW EC3.4.16.5.

XX

OS Aspergillus niger.

XX

```

PN WO200268623-A2.
XX 06-SEP-2002.
XX 22-FEB-2002; 2002WO-EP01984.
XX 23-FEB-2001; 2001EP-0200657.
PR 23-FEB-2001; 2001EP-0200658.
PR 26-FEB-2001; 2001EP-0200660.
PR 26-FEB-2001; 2001EP-0200706.
PR 26-FEB-2001; 2001EP-0200707.
PR 26-FEB-2001; 2001EP-0200708.
PR 26-FEB-2001; 2001EP-0200719.
PR 28-MAR-2001; 2001EP-0000075.
PR 28-MAR-2001; 2001EP-0000078.
PR 28-MAR-2001; 2001EP-0000080.
PR 28-MAR-2001; 2001EP-0000087.
PR 28-MAR-2001; 2001EP-0000088.
PR 21-MAY-2001; 2001EP-0000156.
PR 21-MAY-2001; 2001EP-0000159.
PR 21-MAY-2001; 2001EP-0000160.
PR 21-MAY-2001; 2001EP-0000162.
PR 21-MAY-2001; 2001EP-0000165.
PR 21-MAY-2001; 2001EP-0000166.
PR 21-MAY-2001; 2001EP-0000168.
PR 21-JUN-2001; 2001EP-0000240.
PR 21-JUN-2001; 2001EP-0000242.
PR 21-JUN-2001; 2001EP-0000244.
PR 21-JUN-2001; 2001EP-0000246.
PR 12-JUL-2001; 2001EP-0000280.
PR 12-JUL-2001; 2001EP-0000285.
PR 30-JUL-2001; 2001EP-0000323.
PR 30-JUL-2001; 2001EP-0000327.
PR 02-AUG-2001; 2001EP-0000341.
PR 02-AUG-2001; 2001EP-0000342.
PR 02-AUG-2001; 2001EP-0000343.
PR 02-AUG-2001; 2001EP-0000344.
PR 09-AUG-2001; 2001EP-0000357.
PR 16-AUG-2001; 2001EP-0000374.
PR 16-AUG-2001; 2001EP-0000377.
PR 20-SEP-2001; 2001EP-0000478.
PR 20-SEP-2001; 2001EP-0000483.
PR 22-OCT-2001; 2001EP-0000552.
PR 22-OCT-2001; 2001EP-0000553.
PR 22-OCT-2001; 2001EP-0000554.
PR 22-OCT-2001; 2001EP-0000556.
PR 22-OCT-2001; 2001EP-0000557.
PR 15-NOV-2001; 2001EP-0000558.
PR 21-DEC-2001; 2001EP-00005117.
XX (STAM ) DSM NV.
XX Edens L, Van Dijk AA, Krubasik P, Albermann K, Stock A, Kimpel B;
PI Klugbauer S, Wagner C, Fritz A, Von Gustedt W, Heinrich O;
PI Maier D, Spreafico F, Folkers U, Hopper S, Kemmner W, Tan P;
PI Stiebler J, Albarg R;
XX WPI; 2002-723203/78.
DR P-PSDB; ABZ78210, ABZ78267.
XX
XX Novel isolated protease polypeptide useful in laboratory, clinical,
PT pharmaceutical, chemical, diagnostic, personal care and industrial
PT applications
XX
XX Claim 13; Page 329-331; 394pp; English.
XX
XX The invention relates to a novel isolated protease polypeptide. A
CC polypeptide or polynucleotide of the invention is useful for diagnosing a
CC fungal infection such as aspergillosis, or as a query sequence to perform
CC a search against public databases. A polypeptide of the invention is
CC useful in a selected number of industrial or pharmaceutical processes, in
CC laboratory or clinical processes, in food industry (baking, brewing,

```

```

CC cheese manufacturing, meat tenderising), in tanning industry and in the
CC manufacture of biological detergents. A polypeptide may also be useful
CC for improving protein solubility, extraction yields, viscosity or taste,
CC texture, nutritional value, minimising of antigenicity or
CC anti-nutritional factors, colour or functionality as well as processing
CC aspects like filterability of the proteinaceous raw material. The
CC sequences shown in ABR38869 represent the A. niger proteases of
CC the invention.
XX
SQ Sequence 455 AA;

```

```

Query Match 12.6%; Score 365; DB 23; Length 455;
Best Local Similarity 27.3%; Pred. No. 3.1e-23;
Matches 135; Conservative 68; Mismatches 197; Indels 94; Gaps 22;

```

```

QY 3 GSTPASVGRKQLPKNPTGVKTLTANNVTIRYKEFGAEGVCTTPEGVKSYSYGVDTSPES 62
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 25 GADPTVFRS--PHSPA-----PSIRIQEQN-DSICDARS--PQFTGLWLDIGPK- 68
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 63 HTFFWFFFAHNPETAPITLWNLGGGSDSLGLFEELGPGCHVN--STFDDYINPHSWNE 120
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 69 HLFWYFESQNDPFHDPLTLWMTGGPGDSSMIGLEFEEVGPGRINEFGNGTIDH-NEWAWTK 127
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 121 VSNLLFLSOPLGVGFSYSDTVDGSLNPVTGVVENSFAGVQGRYPTIDAILDITNLAAE 180
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 128 NSSLLFVDQPDVGVFSYID-----EGYE-----LPHDSREAAV 160
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 181 AAWEYLQGLSLGSLDSRVQSKDFSLWTESYGGHYGPAFFNHFHFEQNERIANGSVNGVQ 240
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 161 DMHRELRLFIS--EIFPHKQFLPVHLSGESVAGRYIPYLATQILEQNELYKDSP----R 213
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 241 LNFNSLGIINGLIIDEAIOAPYYPFAVANNYGIKA--VNETVYNYMKTANQMPNGCDLI 298
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 214 IPLKSLGVNGFMSPKDATFGYWEITCTNSGVPSPIETNCRDLM--AANNPH-CMDLY 270
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 299 STCKQTNTALADYALCAEATNMCEDNVEGYPYAFAGRVYDIRHPYDDPTPP----- 351
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 271 DICIQH-----SDPAICHAASVCYDSVVGMLAKLLLR-----MTVTAPCEIDEMC 317
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 352 ----SYNKFLLAKDSVMDAIGVNINYTQSN---NDVYVAFQCTGDFVWPNFIEDLEEL-- 402
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 318 YIEAALIERYLNSPSVWEALSPPOQVTEYKFAVATSIDAFQASAD----GMVSSSKQIAF 373
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 403 -LALPVRVSLYIGDADYTCNNWFGGOAVSLAANYSOAQRFSAGYIPLKXNGVEYGE--TRE 460
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 374 LLANNVDFLAYQGNLDLACNTAGNLRWANSLSKWQTEFTAKPLLPWEIQ-VSVGEGTDE 432
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 461 YGNFSEFTRVYEAGH 474
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 433 TSRAFAVTVDNAGH 446
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Search completed: September 16, 2003, 22:47:51
Job time : 76 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: September 16, 2003, 22:54:28 : Search time 121 seconds
(without alignments)
1958.866 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555
Perfect score: 2887
Sequence: 1 LPGSTPASVGRRLPKNFTG.....HTQSSVPLPTATMSVGVMA 537

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_P/US09712338/runat_16092003_144915_19165/app.query.fasta_1.711
-DB=Issued Patents_NA -QPM=fastcap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -STAXI=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338_SCGN_1_103_runat_16092003_144915_19165 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA :
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq : *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq : *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq : *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq : *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq : *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2883	99.9	1662	3	US-08-943-714-1
2	623	21.6	2002	1	US-08-309-341-3
3	623	21.6	2002	1	US-08-608-267-3
4	623	21.6	2002	1	US-08-608-452-3
5	623	21.6	2002	1	US-08-608-224-3
6	623	21.6	2002	2	US-08-967-149-3
7	594	20.6	2068	1	US-08-309-341-1
8	594	20.6	2068	1	US-08-608-267-1
9	594	20.6	2068	1	US-08-608-452-1
10	594	20.6	2068	1	US-08-608-224-1
11	594	20.6	2068	2	US-08-967-149-1
12	568.5	19.7	2503	1	US-09-640-305-3

13	568.5	19.7	2503	1	US-08-360-673-3	Sequence 3, Appli
14	523.5	18.1	2632	2	US-08-899-524-32	Sequence 32, Appl
15	523.5	18.1	2632	3	US-08-329-892B-32	Sequence 32, Appl
16	361	12.5	1551	2	US-08-828-488-4	Sequence 4, Appli
17	361	12.5	1551	4	US-09-299-689A-4	Sequence 4, Appli
18	348.5	12.1	1670	2	US-08-828-488-2	Sequence 2, Appli
19	348.5	12.1	1670	4	US-09-299-689A-2	Sequence 2, Appli
20	335	11.6	1814	4	US-09-702-705-319	Sequence 319, App
21	335	11.6	1814	4	US-09-736-457-319	Sequence 319, App
22	246.5	8.5	1604	3	US-08-665-966-9	Sequence 9, Appli
23	246.5	8.5	1604	3	US-09-041-780-9	Sequence 9, Appli
24	169.5	5.9	1263	2	US-08-828-488-6	Sequence 6, Appli
25	169.5	5.9	1263	4	US-09-299-689A-6	Sequence 6, Appli
26	163	5.6	1960	4	US-09-280-116-133	Sequence 133, Appl
27	161.5	5.6	497	4	US-09-643-597-190	Sequence 190, App
28	161.5	5.6	497	4	US-09-480-884A-190	Sequence 190, App
29	161.5	5.6	497	4	US-09-542-615A-190	Sequence 190, App
30	161.5	5.6	497	4	US-09-606-421B-190	Sequence 190, App
31	154	5.3	357	4	US-09-280-116-131	Sequence 131, App
32	154	5.3	607	4	US-09-280-116-130	Sequence 130, App
33	130.5	4.5	522	3	US-08-998-416-160	Sequence 160, App
34	127.5	4.4	3279	4	US-09-634-238-44	Sequence 44, Appl
35	127.5	4.4	7818	4	US-09-634-238-14	Sequence 14, Appl
36	126	4.4	1155	2	US-08-387-942C-18	Sequence 18, Appl
37	126	4.4	12588	2	US-08-387-942C-1	Sequence 1, Appli
38	122.5	4.2	2948	4	US-09-865-621A-6	Sequence 6, Appli
39	122.5	4.2	3427	4	US-09-865-621A-8	Sequence 8, Appli
40	122.5	4.2	3518	4	US-09-865-621A-5	Sequence 5, Appli
41	122.5	4.2	3979	4	US-09-865-621A-4	Sequence 4, Appli
42	122.5	4.2	4810	4	US-09-865-621A-3	Sequence 3, Appli
43	122.5	4.2	5163	4	US-09-865-621A-7	Sequence 7, Appli
44	122.5	4.2	5509	4	US-09-865-621A-1	Sequence 1, Appli
45	122	4.2	1962	4	US-09-252-991A-4847	Sequence 4847, Ap

ALIGNMENTS

RESULT 1

US-08-943-714-1
; Sequence 1, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION:
; APPLICANT: Blinkovsky, Alexander
; APPLICANT: Berka, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Dammann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 61875780 No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,714
FILING DATE: 103-OCT-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4990.200-US
TELECOMMUNICATION INFORMATION:
102e

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1662
OTHER INFORMATION:
US-08-943-714-1

Alignment Scores:
Pred. No.: 7,61e-308 Length: 1662
Score: 2883.00 Matches: 536
Percent similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.86% Indels: 0
DB: 3 Gaps: 0

US-09-712-338-2_COPY_19_555 (1-537) x US-08-943-714-1 (1-1662)

Qy	1	LeuProGlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGly	20
Db	55	CTCCAGGAAGTACACCGGCGCTCGGTGAGAGACAGCTACCAAGAACCCACCGGG	114
Qy	21	VallysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu	40
Db	115	GTCAGAGACTCTACACCGCAACAATGTCACCATCGGTACAAAGAACCCCGGGCAGAG	174
Qy	41	GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro	60
Db	175	GGCGTCTGCGAGACTACCCCGGGTCTCAATCTCTACTGTGATATGTCGACACTCTCC	234
Qy	61	GluSerHisThrPheThrPheGluAlaArgHisAsnProGluThrAlaProIle	80
Db	235	CAGTCCCNATCTCTCTGTTCTTCGAGCCGACATACCCAGAACTGCACCTATC	294
Qy	81	ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu	100
Db	295	ACATTTGGTTGAATGGTGGCCCTGGAAGGATCTTTTGATCGGTCTCTCGAAGAGTTG	354
Qy	101	GlyProCysHisValAsnSerThrPheAspAspIleAsnProHisSerTrpAsnGlu	120
Db	355	GGCCCTTGCCATGCAATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGAAACGAG	414
Qy	121	ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr	140
Db	415	GTCTCCAAATTTACTATCTCTGCCAGCATTTGGGAGTCCGGCTTTTCATATAGTATAGG	474
Qy	141	ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal	160
Db	475	GTGTATGGTGCATTAACCCCTGAATCGGGTGGTGGAAATTCGAGCTTTTCAGGAGTT	534
Qy	161	GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGlu	180
Db	535	CAGGCGCGGTACCCCAACATTGATGCTCTGATGATGATGATGATGATGATGATGATGATG	594
Qy	181	AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal	200
Db	595	GCCGCTTGGAGATCTCGAAGATTCCTTACTGGACATACCTAGCTTGGACTTAGGGTG	654
Qy	201	GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhe	220
Db	655	CAGTCTAAGAGCTTCAGTCTATGGAGCGGAGAGCTATGGAGGCACTATGGTCTCTCATTC	714
Qy	221	PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln	240
Db	715	TTCAATCATTTTTTACGACGAGATGAGAGAAATTCGACCGGTAGTGTAAATGCTTCAG	774

Qy	241	LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGleAlaPro	260
Db	775	CTTAATTTCAACTCTCTGGGAATTAATTAACGGCATCATCGACGAGGCGATCCAGGCCCT	834
Qy	261	TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal	280
Db	835	TACTACCTGAATTCGGTGTGAACAATACCTACGGTATCAAGGTATCAAGGTATCAAGGTAT	894
Qy	281	TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr	300
Db	895	TACAACATCACTGAAGTTTGCACCAATGCTTGCAGGATTTGCAGGATTTGCAGGATTT	954
Qy	301	CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuLysAlaGluAlaThrAsn	320
Db	955	TGCAACACAGACAAACCGCACCGCATTAGCTGACTACGGCTCTGCGCGAAGCCACCAAC	1014
Qy	321	MetCysArgAspAsnValGluGlyProTyrThrAlaPheAlaGlyArgGlyValTyrAsp	340
Db	1015	ATGTGACGGGCAATTTGAGGGGCACTACACGCTTTCCTGCTGCTGCTGCTGCTGCTGCT	1074
Qy	341	IleArgHisProTyrAspProThrProSerTyrTyrAsnLysPheLeuAlaLys	360
Db	1075	ATTCGGCATCATATGATGACCCGACTCCGCCAAGTTATACAAACAATTTCTGCAAAAG	1134
Qy	361	AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal	380
Db	1135	GACTCTGTCATGGACGCTATCGGCTCAACATCACTACCCACTACCCACTACCAATTAATGCTC	1194
Qy	381	TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu	400
Db	1195	TACTACGCTTCCAGCAACAGGCGACTTTGCTGCGCCCACTTCTATCGAGACCTCGAG	1254
Qy	401	GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys	420
Db	1255	GAGATCTTCTCTCCGCTGCTCTCCCTCATCTATGGCGACGCGGATTTACATCTGC	1314
Qy	421	AsnTrpPheGlyGlyGluAlaValSerLeuAlaAlaAsnTyrSerGlnAlaGlnPhe	440
Db	1315	AACTGGTTCGGCGGTGAGCGCTTTCCTCTGCTGCGAATCTCTCCCAAGCCCGCAGTTC	1374
Qy	441	ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu	460
Db	1375	CGAAGCGGAGGTATACGCGCTGAAAGTCAACGGCTGAGTATGGGAAACTCCGAG	1434
Qy	461	TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTrpGln	480
Db	1435	TATGTAATTTCTCTCTTCTATGAGCGAGGCAATGAAGTCCCATATACCTACAG	1494
Qy	481	ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGly	500
Db	1495	CCCATCGCTCTCTGCAATTTTAAACCGGACTATCTTCGTTGGGATATCGCAGAGGC	1554
Qy	501	GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln	520
Db	1555	CAGAAGAGATCTGCGCCAGCTACAGCAATGGAACGCTACAGCTACAGCTACAGCTACAG	1614
Qy	521	SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet	536
Db	1615	TCGTCGCTGCGCTGCTACGCTACCAAGCATGTCAGTGTGTTGATG	1662

RESULT 2
US-08-309-341-3
Sequence 3, Application US/08309341
Patent No. 5594119
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
APPLICANT: Thompson, Sheryl Ann
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 55941190 No. 5594119th disk of No. 5594119th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400

CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA: US/08/309,341
 APPLICATION NUMBER: US/08/309,341
 FILING DATE: 16-SEP-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lowney, Karen A.
 REGISTRATION NUMBER: 31,274
 REFERENCE/DOCKET NUMBER: 4247.000-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 867 0298
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2002 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Aspergillus niger
 FEATURE:
 NAME/KEY: intron
 LOCATION: 349..411
 NAME/KEY: CDS
 LOCATION: join (348..412)
 US-08-309-341-3

Alignment Scores:

Pred. No.: 1,29e-58
 Score: 623.00
 Percent Similarity: 48.25%
 Best Local Similarity: 31.96%
 Query Match: 21.58%
 DB: 1

US-09-712-338-2_COPY_19_555 (1-537) x US-08-309-341-3 (1-2002)

QY 27 AlaAsnValThrThrLeArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
 DB 533 GCCTATGATCTCAGGTCAGAGAACCGATCTCTTGGCATC-----GAC 583
 QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 65
 DB 584 CTGGCGCTAAGACGATACCGGTTATCTCGATGACACGAGAACGACGACGATCTGTC 643
 QY 66 PheTrpPhePheGluAlaArgHisAsuProGluThrAlaProIleThrLeuTrpLeuAsn 85
 DB 644 TACTGGTTCTTCAGTCTCGCATGATCCCGGAGAAATGACCCCTGTTCTGTCGGCTGAAC 703
 QY 86 GlyGlyProGlySerAspSerLeuLeuGlyLeuPheGluGluLeuGlyProCysHisVal 105
 DB 704 GTFGGCCCTGGATGCTCTCCCTACCGGTCCTTTTCATGAGCTCGGCCCTAGCAGATC 763
 QY 106 AsnSerThrPheAspAspThrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeu 125
 DB 764 ACAAGAGATCATCCGCGGTCTACACGACTACGCTTGGAACTCCACGCGCTCGGTGATC 823
 QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
 DB 824 TTCCTTGACCGCTGTCAGGTCGGTTACTCTACGACCACTCTGCT----- 871
 QY 146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165

DB 871 ----- 871
 QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaThrPrgluIle 185
 DB 872 -----GICAGGACACCGCTGCTGCTGCGCAGGAGCTCTATGCCCTTG 913
 QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205
 DB 914 CTTACCCCTCTTCTTCAACAATTCCTC-----GAGTATGCCAAGCAGGACTTC 961
 QY 206 SerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyr 225
 DB 962 CACATGCCGCTGAATCTATGCTGGTCACTATATCCCGCTCTTGTCTCGGAGATTPTG 1021
 QY 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245
 DB 1022 TCTCACAAGAGCGC-----AACATCAACCTGCAGTCC 1054
 QY 246 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 265
 DB 1055 GTCCTTATGGAACCGGTCTCACCGACGGTCTCACTCAGTACGAGTACTACCGTCCCATG 1114
 QY 266 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 285
 DB 1115 GCCTGTGTGACGGT---GGTACCAGCTGCTCTGGACGAG---GGCTCTGCCAGGCC 1168
 QY 286 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 305
 DB 1169 ATGACACAACCGCTCTCTCGC---TGGCAGTCTATGATTAGTCTTGTATAGTTCGCG 1225
 QY 306 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 325
 DB 1226 AGCGCT-----TGGGTTGTGTCGCCGCTCCATCTACTACTAACAAGGCC 1270
 QY 326 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 345
 DB 1271 CTCCTTCCCTTACCAAGCG---ACCGGACAGACGCTCAGTATGTCGTGGTAAGTGC 1327
 QY 346 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360
 DB 1328 GAGGATAGCTCCAACTCTGCTACTGCGCATGGCTACGTCAGTCAGCGACTACTCTGAACAAG 1387
 QY 361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 379
 DB 1388 ACCGAGTCAATGAGGCTGTTGGCGCTGAGGTCACGCTACGACTCGTCAACTTTCAC 1447
 QY 380 ValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 397
 DB 1448 ATCAACGCGAACCTCTCTTCCACGGTGACTGGATGAGCGCTACCACTCGCTCGTTCGG 1507
 QY 398 AspLeuGluGluLeuLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416
 DB 1508 GGACTCTCTGGAG-----CAGATCCCTGTC-----CTGATCTACGCTGGTGGACGCC 1552
 QY 417 AspTyrIleCysAsnTrpPheGlyGlyGluAlaValSerLeuAlaAlaAsnTyrSerGln 436
 DB 1553 GATTTATCTGCAACTGGCTGGCGCAACAGCGCTGGAGCCCTTGATGGTGGCCCGGA 1612
 QY 437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451
 DB 1613 CAGGCTGAATATGCTTCCCTTAAGCTGGAGGACCTGCTCGTGGTTCGAGAAATGACACAAG 1672
 QY 452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471
 DB 1673 GCGAAGAAGATCGCCAGGTCAGTCCCATGGCACTTCCACTTCACTGCTCTCTATGSC 1732
 QY 472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491
 DB 1733 GTGGGCCCATGTCCTCCGATGGACCAACCGAGCTCGAGTCTTGAAATCTTCAACCGCTGG 1792
 QY 492 IlePheGly---Trp 495
 DB 1793 TTGGGAGGTCATGG 1807


```

QY 437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451
DB 1613 CAGGCTGAATATCCCTCCGCTAGCTAGCGAGGAGCTGGTGTGCGAGAAATGAGCACAAG 1672
QY 452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471
DB 1673 GCAAGAACATCGCCAGGTCAGAGTCCCATGCGCAAGTTCACCTTCATGCGTCTCTATGGC 1732
QY 472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491
DB 1733 GTGGCCCATGTCGCCGATGGACCAACCGAGTCGAGTCTGAATTCCTCAACCGCTGG 1792
QY 492 IlePheGly---Trp 495
DB 1793 TTGGGAGGTGAATGG 1807

RESULT 4
US-08-608-452-3
; Sequence 3, Application US/08608452
; Patent No. 5693510
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56935100 No. 5693510disk of No. 5693510th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,452
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join (348..412)
; US-08-608-452-3

Alignment Scores:
Pred. No.: 1,29e-58 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79

```

```

Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.58% Indels: 76
DB: 1 Gaps: 18

US-09-712-338-2_COPY_19_555 (1-537) x US-08-608-452-3 (1-2002)

QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
DB 533 GCCTAGTAUCTCAGGTCACAGAACGCAICCTAGCTCTCTTGCAATC-----GAC 583
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 65
DB 584 CTTGGGTAAAGCAGTACACCGGTATCTCGATCACACGACGACGACACGACATCTTTC 643
QY 66 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuThrLeuAsn 85
DB 644 TACTGTCTCTCGAGTCTCGCAATGACCCGAGAAATGACCCCTGTGTCTGTGCTGAAC 703
QY 86 GlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyProCysHisVal 105
DB 704 GTGGCCCTCGATGCTCTTCCCTACCGGTCTTTCATGAGCTCGGCCCTAGCAGCATC 763
QY 106 AsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeu 125
DB 764 AACAAAGATCCAGCGGTCTACACGACTACGCTTGGAACCTCCACGCGTCCGTGAIC 823
QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
DB 824 TTCCTTGACACGCTCTCAACGTCGTTACTCTTACAGCAACTCTGCT-----871
QY 146 AsnProValThrGlyValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165
DB 871 -----871
QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaIleTrpGluIle 185
DB 872 -----GTCAGCGACACCGTTGCTCTGCGAAGGAGCTATCCCTTG 913
QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205
DB 914 CTTACCCCTCTCTTCAACAAATCCCC-----GAGTATGCCAAGCAGGACTTC 961
QY 206 SerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyr 225
DB 962 CACATTGCCGCGTAATCTTATGCTGTCATATATCCCGCTCTTCTCGGAGATTTG 1021
QY 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245
DB 1022 TCACACAAGAGCGC-----AACATCAACCTGCAGTCC 1054
QY 246 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPhe 265
DB 1055 GTCTTATTGGCAACGGTCTCACCGCGTCTCAGTCAGTACGAGTACTACCGTCCCATG 1114
QY 266 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnIleMetLys 285
DB 1115 GCCTGTGTGAGCGT---GTTACCCAGCTGTCTGGACGAG---GGCTCCGCCAGGCC 1168
QY 286 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 305
DB 1169 ATGGACACGCGCTTCCCTCGC---TGCAGCTATGATTGAGTCTGCTATAGTCCGAG 1225
QY 306 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsn 325
DB 1226 AGCGCT-----TGGTGTGTGTCGCGCTCCATCTACTGTAAACAACGCC 1270
QY 326 ValGluGlyProTyrThrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 345
DB 1271 CTCCTTGCCCTTACCAGCGC---ACCGAGACAGACGCTCTTACGATCTTCGTGTAAGTGC 1327
QY 346 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360
DB 1328 GAGGATAGCTCCAACTCTGCTACTCGGCCATCGGCTAGCTCAGCGACTACCTGACAACAG 1387

```

```
QY 361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsnAsp 379
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1388 ACCGAGTATGAGGCTGTGGCTGAGGTCAAGCGTACGACTCGTCCACTTTGAC 1447
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 380 ValTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIleGlu 397
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1448 ATCAACCGCAACTTCTCTTCCACGGTGACTGGATGAAGCCCTACCAACCGTCTCGTTCCG 1507
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 398 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1508 GGACTCTCTGAG-----CAGATCCCTGTC-----CTGATCTACGCTGTGTGAGCC 1552
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 417 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGln 436
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1553 GATTTTCATCTGCNACTGGCTGGCAACAGGCTCGACTCAAGCCCTTGAGTGCCCGGA 1612
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1613 CAGGCTGAATATGCCCTCCGCTAAGCTGGAGACTGGTGGTGGAGAAATGAGCACAAG 1672
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1673 GGCAAGAAGATCGCCGAGTCAAGTCCCATGGCAACTTCACCTTCATGCGTCTCTAIGGC 1732
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1733 GGTGGCCACATGTCCTCCGATGGACCAACCCGAGTCGAGTCTTGAAATCTTCAACCGCTGG 1792
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 492 IlePheGly---Trp 495
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1793 TTGGGAGGTGAATGG 1807
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 5
US-08-608-224-3
; Sequence 3, Application US/08608224
; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 57053760 No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (348..412)
US-08-608-224-3

Alignment Scores:
Pred. No.: 1,29e-58 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.58% Indels: 76
DB: 1 Gaps: 18

US-09-712-338-2_COPY_19_555 (1-537) x US-08-608-224-3 (1-2002)
QY 27 AlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 533 GCCTATGATCTCAGGGTCAAGAGACGATCTCTAGCTCTCTTGGCATC-----GAC 583
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 65
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 584 CTGGCGTAAAGCAGTACACCGGTATCTCGATGACACGAGACGACACGATCTGTC 643
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 66 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 85
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 644 TACTGTTCTTCGAGTCTCGAATGACCCGAGAAATGACCTGTGTGTTCTGTGGTGAAC 703
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 86 GlyClyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisVal 105
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 704 GGTGGCCCTGGATGCTCTCCCTCACCCTGCTTTTCATGGAGCTCGGCCCTAGCAGCATC 763
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 106 AsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeu 125
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 764 AACAGAAGATCCAGCCGGTCTACACGACTACGCTTGGAACTCCAACGGTCCGTGATC 823
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 824 TTCTTACCAGCGCTGCAACGCTGGTTACTTCTTACAGCAACTCTGCT----- 871
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 871 ----- 871
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaIleTrpGluIle 185
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 872 -----GTCAGCGACACCGTTCCTGCTGCAAGGAGCGTCTATGCTTG 913
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 914 CTTACCTCTCTTCAAAACAATTCCCC-----GAGTATGCCAAGCAGGACTTC 961
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 206 SerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyr 225
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 962 CACATTGCCGTGAATCTCTATGCTGGTCACTATATCCCGCTCTTGGCTGGAGATTGG 1021
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1022 TCTCACAAGAAGCGC-----AACATCAACCTCGACTGC 1054
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 246 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpProGluPhe 265
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1055 GTTCTTATGGCAACGCTCTCACGACGGTCTCACTCAGTACGAGTACTACCGTCCCATG 1114
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 266 AlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLys 285
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1115 GCCTGTGTGACGGT---GGTTACCCAGCTGCTGTGGACGAG---GGCTCTCGCCAGGCC 1168
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

QY      286 PheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsn 305
Db      1169 ATGGACACGCCCTCTCTCGC---TCCAGTGTATGATGTGCTGTATGCTATGCTCCGAG 1225
QY      306 ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsn 325
Db      1226 AGCGCT-----TGGGTTTGTGTCCCGCTCCCTACCTACCTACCAACGCC 1270
QY      326 ValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyr 345
Db      1271 CTCCTTGCCCTTACACGCC---ACCGACAGACAGCTGTACGATGTGCTGTAAGTGC 1327
QY      346 AspAspPro-----ThrProSerTyrTyrAsnLysPheLeuAlaLys 360
Db      1328 GAGGATAGTCCCACTCTGCTACCTGGCGCATGGCTAGCTAGCAGTACCTGACACAG 1387
QY      361 AspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsp 379
Db      1388 ACCGAGTGTATGAGGCTGTGGCGCTGAGGTCAACGGCTACGACTCGTGCAACTTIGAC 1447
QY      380 ValTyrTyrAlaPheGlnThrGlyAspPheValPro-----AsnPheIleGlu 397
Db      1448 ATCAACCGCACTCTCTTCACCGGTGACTGGATGAAGCCCTACCAACGCTCTCGTTCCG 1507
QY      398 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAspAla 416
Db      1508 GGACTCTCTGAG---CAGATCCTCTGTC-----CTGATCTACCGTGTGACGCC 1552
QY      417 AspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGln 436
Db      1553 GATTCACTGCACTGGCTGGCGCAACAGCGCTGGACTGAAGCCCTTGGTGGCCCGCA 1612
QY      437 AlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----Asn 451
Db      1613 CAGGTGTGAATGCTCCGCTGAAGCTGGAGGACCTGGTGTGCTGCGAGATGACACAA 1672
QY      452 GlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGlu 471
Db      1673 GCACAAGATGCGCCAGCTGAGTCCAGTCCCATGGCACTTCACCTCATGCTCTAAGGC 1732
QY      472 AlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThr 491
Db      1733 GGTGGCCCATGCTCCCGATGACCAACCCGAGTGGAGTCTTGAATCTTCAACCGCTGG 1792
QY      492 IlePheGly---Trp 495
Db      1793 TTGGAGGTGAATGG 1807

```

RESULT 6

```

US-08-967-149-3
; Sequence 3, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5939305 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CORRESPONDENCE ADDRESS:
; APPLICATION NUMBER: US/08/967,149

```

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; AUTOREY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 349..411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join (348..412)
; US-08-967-149-3

```

```

Alignment Scores:
Pred. No.: 1,29e-58 Length: 2002
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.58% Indels: 76
DB: 2 Gaps: 18

```

```

US-09-712-338-2_COPY_19_555 (1-537) x US-08-967-149-3 (1-2002)
QY      27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db      533 GCCTATGATCTCAGGCTCAAGAACCGATCTCTCTGTGGCATC-----GAC 583
QY      47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerPro---GluSerHisThrPhe 65
Db      584 CCTGGCGTAAAGCAGTACACCGGTTATCTCGATGACACGAGAACGACGACATCTGTC 643
QY      66 PheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsn 85
Db      644 TACTGGTCTCTCGAGTCTCGCAATGACCCCGAGATGACCCCTGTTCTGTGGCTGAAC 703
QY      86 GlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisVal 105
Db      704 GTTGCCCTGGATGCTCTCTCCCTACCGGTCTTTCAIGAGCTCGGCCCTAGACGATC 763
QY      106 AsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 125
Db      764 AACAAAGAGATCCAGCGGCTCTACCAACGACTACGCTTGGAACTCCCAACGCGTCCG 823
QY      126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
Db      824 TTCCTTGACACGCTGTCAACGTCGGTTACTTACTTACAGCAACTCTGCT----- 871
QY      146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165
Db      871 ----- 871
QY      166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 185
Db      872 -----GTCAGGACACCGCTGTGCTGTGGCAAGGACGCTCTAIGCCTGTG 913

```



```

Qy 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470
    ||| ::: ||||| ::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1656 ACGGCAAGAGATGTGCCAGGTAACTCCCATGGCAACCTTCACTTCGCTCTAT 1715
Qy 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 GGTGGTGGCCACATGGTCCGATGGACGACCGCCAGTCTCGAGTCTTCAACCGC 1775
Qy 491 ThrIlePheGly---Trp 495
    ::: ||| |||
Db 1776 TGGTTGGGAGGTGAATGG 1793

RESULT 8
US-08-608-267-1
; Sequence 1, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56886630 No. 5688663disk of No. 5688663th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (571..633)
US-08-608-267-1

Alignment Scores:
Pred. No.: 2,14e-55 Length: 2068
Score: 594.00 Matches: 155
Percent Similarity: 46.44% Conservative: 80
Best Local Similarity: 30.63% Mismatches: 174
Query Match: 20.57% Indels: 98
DB: 1 Gaps: 18

```

US-09-712-338-2_COPY_19_555 (1-537) x US-08-608-267-1 (1-2068)


```
/ MOLECULE TYPE: Genomic DNA
/ ORIGINAL SOURCE:
/ ORGANISM: Aspergillus niger
/ FEATURE:
/ NAME/KEY: intron
/ LOCATION: 572..632
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: join (571..633)
US-08-608-452-1

Alignment Scores:
Pred. No.: 2,14e-55
Score: 594.00
Percent Similarity: 46.44%
Best Local Similarity: 30.63%
Query Match: 20.57%
DB: 1
Gaps: 18

US-09-712-338-2_COPY_19_555 (1-537) x US-08-608-452-1 (1-2068)

QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db 457 GCCTATGATCTCAGGTCAGAGAGACCGATCTCGCTCTCTTGGCATC-----GAC 507
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu----- 61
Db 508 CCGCGCGTGAAGCAGTACCGGTATCTCGATGACACAGCA-GAATGATAGCAATTGTT 566
QY 62 -----SerHisThrPhePhe----- 66
Db 567 CTACGTAAAGCACACCTTGGTTCAGATCACCGCTTTTATATGCTCGATATCAACGCA 626
QY 67 -----TrpPhePheGluAlaArgHisAsnProGluThrAlaProIleIhrLeuTrpLeu 84
Db 627 ACTTAGTGGTCTCGAGTCTCGCAATGACCCGAGAGATGATCCCGTGTCTCTGTGGCTG 686
QY 85 AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHis 104
Db 687 AACGGTGGCCCTGGGTCTCTCCCTACCGGCTCTTCTATGAGCTTGGCTTGGCTTGGCTG 746
QY 105 ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 124
Db 747 ATCAACAAGAAAGATCCACCGCGGTCTACAATGACTACGCTTGAACCTCAACGCGTCCG 806
QY 125 LeuPheLeuSerGlnProGluValGlyPheSerTyrSerAspThrValAspGlySer 144
Db 807 ATCTTCTTCCAGCAGCTGTCAATGTGGTACTCTCTACAGTAACTCTGCT----- 857
QY 145 IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr 164
Db 857 ----- 857
QY 165 ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaIleTrpGlu 184
Db 858 -----GTCACGACACAGCTCGCTGCTGCGAGGAGGAGCTATGCC 896
QY 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 204
Db 897 TTGCTTACCTCTCTTCAACAATTCCTCC-----GAGTATGCTAAGCAGGAC 944
QY 205 PheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPhe 224
Db 945 TTCCATTGCGCGTGAATCTTATGTGCTACATATATCCCGCTTCTCGGATCGGAGATC 1004
QY 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244
Db 1005 CTGCTCTCAAGAAGAGCG-----ACATCAACCTTCGAC 1037
QY 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpProGlu 264
Db 1038 TCGGTTCTTATGGCAACGCTCTCACCGGGAATACACCCAGTACGAGTACTACCGTCC 1097
```

```
QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
Db 1098 ATGCCICGGGTGACGCGGCTTACCCAGCTCTCTTGACGAGAGCTCTGCCAGTCCATG 1157
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db 1158 GAC-----AACGCTCTCTCCGCG-----TGCCAGTCTATGATGAGTCTTGACAGTTC 1208
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaIhrAsnMetCysArgAsp 324
Db 1209 GAGAGCGCT-----TGGGTGTGTCGCGCTCCATCTACTGTAAACAC 1253
QY 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344
Db 1254 GCCCTCTTIGCCCTTACACGCG-----ACTGGCAGAACGCTATGATGCTCGGTGAAG 1310
QY 345 TyrAspAspPro-----ThrProSerTyrTyrAspLysPheLeuAla 359
Db 1311 IGCAGGATAGCTCTAACCTTTGCTACTCGGCTATGCGCIACGTCAGGACTACCTGAAC 1370
QY 360 LysAspSerValMetAspAlaIleGlyValAsnIleAsn-----TyrThrGlnSerAsnAsn 378
Db 1371 AAGCCGAAGTCAATCGAGGCTGTGGCGCTCAGGTCAACGGCTACGACTCGTGAACATT 1430
QY 379 AspValTyrTyrAlaPheGlnThrGlyAspPheValTrpPro-----AsnPheIle 396
Db 1431 GACATCAACCCCACTTCTCTTCCAGCTGACTGGATGAAGCCCTACACCGCTCGTT 1490
QY 397 GluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleIleIleIleIle 415
Db 1491 CCGGACTCTGGAG-----CAGATCCCTGTC-----TGATCTATGCGCGTGA 1535
QY 416 AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnIleTrpSer 435
Db 1536 GCTGATTTCATTTCACTGGCTGGCAACAAGCCCTGGACTGAAGCCCTGAGTGGCC 1595
QY 436 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal----- 450
Db 1596 GGACAGCTGAATATGCTCGCTGAGCTGGAGGATCTGGTCAITGTGACATGAGCAC 1655
QY 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470
Db 1656 ACGGCAAGAGATGCGCAGGTAACTCCATGGCACTTCACTTCACTGCGTCTAT 1715
QY 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490
Db 1716 GGTGTGGCACATGTCGCGAIGGACCGAGTCCGAGTCTCGAGTCTTCAACCGC 1775
QY 491 ThrIlePheGly---Trp 495
Db 1776 TGGTGGGAGGTGAATGG 1793

RESULT 10
US-08-608-224-1
; Sequence 1, Application US/08608224
; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5705376 No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
```



```

US-08-967-149-1
; Sequence 1, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5939305 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: intron
; LOCATION: 572..632
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join (571..633)
; US-08-967-149-1

Alignment Scores:
Pred. No.: 2,14e-55 Length: 2068
Score: 594.00 Matches: 155
Percent Similarity: 46.44% Conservative: 80
Best Local Similarity: 30.63% Mismatches: 174
Query Match: 20.57% Indels: 98
DB: 2 Gaps: 18

US-09-712-338-2_COPY_19_555 (1-537) x US-08-967-149-1 (1-2068)
QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db 457 GCGATGATCTCAGGCTCAAGACCGATCTCGTCTCTTGGCATC-----GAC 507
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGlu----- 61
Db 508 CCGCGCGTGAACAGTACACCGGTATCTCGATGACACAGA-GAATGATAAGCATTTGTT 566
QY 62 -----SerHisThrPhePhe----- 66

```

```

Db 567 CTAGTAAGCACACACTTGGTTCAAGATCACCGCTTTTATATGCTCTGGATATCTAACGCA 626
QY 67 -----TrpPheGluAlaAArgHisAsnProGluThrAlaProIleThrLeuTrpLeu 84
Db 627 ACTTAGTGGTTCTTCGAGTCTCGCAATACACCCGAGATGATCCGGTTGTTCTGGCTG 686
QY 85 AsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHis 104
Db 687 AACGGTGGCCCTGGTGTCTCTCCCTCATCGGCTCTCTCATGGAGCTTGGCCCTAGCAGC 746
QY 105 ValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 124
Db 747 ATCAACAAGAAGATCCAGCCCGTCTACATGACTACGCTTGGAACTCCACGCGCTCCGTG 806
QY 125 LeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySer 144
Db 807 ATCTTCCTTGACCAGCCTGTCAATGTCGGTTACTCTCTACATACTAATCTGCT----- 857
QY 145 IleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyr 164
Db 857 ----- 857
QY 165 ProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGlu 184
Db 858 -----GTCAGGCACACGCTCGCTGTCGCAAGGAGGCTCTATGCC 896
QY 185 IleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAsp 204
Db 897 TTGCTTACCCTCTTCTTCAACAATCCCC-----GAGTATGCTAACGAGC 944
QY 205 PheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPhe 224
Db 945 ITCCACATCCCGGTGAATCTTATGCTGTGTCACATATATCCCGCTTCGCTCGAGATC 1004
QY 225 TyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsn 244
Db 1005 CTGTCITCACAGAGCGC-----ACATCAACACTCGAG 1037
QY 245 SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGlu 264
Db 1038 TCCGTTCTCATTTGGCAACGGTCTCACCGACGATACACCCAGTACGAGTACTACCGTCC 1097
QY 265 PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
Db 1098 ATGCGCTGCGGTGACGCGCGTTACCCAGCTGTCTTGGACGAGAGCTCTGCCAGTCCATG 1157
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db 1158 GAC-----AACGCTCTCTCTCGC---TGCAGTCTATGATGAGTCTGTACAGTTCC 1208
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 324
Db 1209 GAGAGCGCT-----TGGGTTGTGTCGCCGCTCCACTCTACTGTAAACAAC 1253
QY 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344
Db 1254 GCCCTCTTGGCCCTTACCAGCGC---ACTGGCAGAACGCTGTATGATGCTGGTGAAG 1310
QY 345 TyrAspAspPro-----ThrProProSerTyrTyrAsnLysPheLeuAla 359
Db 1311 TGGCAGGATAGCTTAACCTTCTACTCGGCTATGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1370
QY 360 LysAspSerValMetAspAlaIleGlyValAsnIleAsn---TyrThrGlnSerAsnAsn 378
Db 1371 AAGCCCGAAGTCAATCAGAGCTGTGGCGTGAAGTCAACGGCTACCGCTACGCTGCTCACTTT 1430
QY 379 AspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpPro-----AsnPheIle 396
Db 1431 GACATCAACCGCAACTCTCTTCCCGGTGACTGATGAAGCCCTACCAACCGCTCGCT 1490
QY 397 GluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr---GlyAsp 415

```

Db 1491 CCGGACTCCTGGAG-----CAGATCCCTGTC-----TTGATCTATGCCGGTGAT 1535
QY 416 AlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnIyrSer 435
Db 1536 GCTGATTTTCATTGCAACAGTGGTGGCGACACAGGCGCTGGACTGAAGCCCTGGAGTGGCC 1595
QY 436 GlnAlaAlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----450
Db 1596 GGACAGCGTGAATATCCCTCCGCTGAGCTGGAGGATCTGGTCAATTCGACATGAGCAC 1655
QY 451 AsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyr 470
Db 1656 ACGGCAAGAAGATTGGCAGGTAAAGTCCCATGCCAACTTCACCTTCATCGCTCTCTAT 1715
QY 471 GluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490
Db 1716 GGTGGTGGCCATGTCCTCCGATGACGACCGAGCCGCGAGTCCGAGTCCGATCTCTCAACCGC 1775
QY 491 ThrIlePheGly---Trp 495
Db 1776 TGGTGGAGGTGAATGG 1793

RESULT 12
US-09-640-305-3
; Sequence 3, Application US/09640305
; Patent No. RE37447
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Yen, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/640,305
; FILING DATE: 16-Aug-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE: 06-FEB-1995
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 387..1862
; OTHER INFORMATION: /product= "K. lactis protease C
; gene"
; /gene= "K1.PRC1"
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-640-305-3
Alignment Scores:
Pred. No.: 1-87e-52 Length: 2503 -
Score: 568.50 Matches: 151
Percent Similarity: 44.65% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 184
Query Match: 19.69% Indels: 85
DB: 1 Gaps: 14
US-09-712-338-2_COPY_19_555 (1-537) x US-09-640-305-3 (1-2503)
QY 27 AlaAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThr 46
Db 570 GCTTATTCATTGAGAAATTAACCTTTGGATCCCAATCTCTTGGCGTT-----617
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePhe 66
Db 618 GATACCGTGAACAATGGTCGGATATTTAGATTACCAGGACTCAACACACTTCTTTAT 677
QY 67 TrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGly 86
Db 678 TGGTTTTTTCAGTCTAGAAATGACCCAGAGAAATGACCCAGTACTACTATGTTAAACGGT 737
QY 87 GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsn 106
Db 738 GGTCTGGCTGTTCTCTCTTCGGGGTCTTTCTTTGAATGGGACCTTCTCTATAGGA 797
QY 107 SerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuPhe 126
Db 798 GCTGATTTGAACCACTTATAACCCCTACTCTTGGATTCACACGCTCTCTGATATTC 857
QY 127 LeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsn 146
Db 858 CTAGATCAGCGCTGTGGTGTGGTCTCTCATACCGTGAC-----896
QY 147 ProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThr 166
Db 896 -----896
QY 167 IleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeu 186
Db 897 -----TCTAAGTGTCTACACAGATGACCTGCCAAGAGCGTTACATATTCTTA 947
QY 187 GlnGlyPheLeuSerGlyLeuProSerIleuAspSerArgValGlnSerLysAspPheSer 206
Db 948 GATTTGTTCTTTGAAGATTCCCTCATTTG-----AGAAATAAGCATTTCCAT 995
QY 207 LeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlu 226
Db 996 ATCTCCGGTGAATCATCAGCGGTCATATTATCCCAAGATGCTCAT-----1043
QY 227 GlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeu 246
Db 1044 -----GAGATTGCTGTAGTGCATGCTGAGGATTCCTCTTCAATCTATCGTCAGTA 1094
QY 247 GlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaProTyrProGluPheAla 266
Db 1095 TTAATTGGAAATGGATTACTGACCCACTGACTCAATACCAATATATAGACCGCATGCC 1154
QY 267 ValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPhe 286
Db 1155 TGTGGTGAAGGT---GGTATCCAGCGGTGTGGAA-----1187
QY 287 AlaAsnGlnMetProAsnGlyCysGlnAsp-----296
Db 1188 -----CCGGAAGATTGCTTACATATGAATAGTAGAATCTACCTCTATGCTATCG 1235

```

QY 297 LeuLeuSerThrCysLysGluThrAsnArgThrAlaLeuAlaAspIyrAlaLeuCysAla 316
DB 1236 CTTGTGGACCGCTGTACAAAGTCCCAT-----TCGTGTTTCTCTGTGTG 1280
QY 317 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 336
DB 1281 TTGGCTGACCGTATPTGTGAACACAGATTACTGGG---GTTTATGAGAAATCAGTAGG 1337
QY 337 GlyValTyrAspIleArg-----HisProTyrAspAspPro-----Thr 349
DB 1338 AACCTTACGATATTAGATCTAAAGTCTCAACGAGGAGGATTCGGTGCCTGTTATCAG 1397
QY 350 ProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVal 369
DB 1398 GAAGAAATTTATCTCTGATCTACTTGAATCAGGAGGAGTTCAGAGAGCTTAGGACT 1457
QY 370 AsnIleAsnTyrThrGlnSer---AsnAsnAspValTyrTyrAlaPheGlnThrGly 388
DB 1458 GATGTGAGTCTTCCCAAGGTTGTAGCTCGGATCTCGGTATCGGTTTCCGATTCCTGCG 1517
QY 389 AspPheValTyrProAsnPheIleGluAspLeuGluGluLeuAlaLeuAlaLeuProValArg 408
DB 1518 GATGACCGAGCCCA---TTCACACAGTACGTCGACAGACTTCTGATCAAGATATCAAT 1574
QY 409 ValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGlyGlyGluAlaVal 428
DB 1575 GTCTTGATATATGACGAGGATAGGATATATTTGTAATTCGCTAGGAAATCTGCTTGG 1634
QY 429 SerLeuAlaAlaAsnTyrSerGlnAlaGlnPheArgSerAlaGlyTyrThrProLeu 448
DB 1635 ACTGAAAAATGGAATGGAGGTATACGAGAGAGGTATAAACAAACAGTTTTCGAGACITGG 1694
QY 449 LysValAsnGlyValGlu-----TyrGlyGluThrArgGluTyrGlyAsnPheSerPhe 466
DB 1695 AAGAGTGAAGAAACAGATGACACCATTTGGCGAACCACCAATCTTATGCGCCGCTAACCTAC 1754
QY 467 ThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGln 486
DB 1755 TTGAGATATATGATGCTGGACACATGTTCTTCACGACCAACCTGAAATTCATTACAA 1814
QY 487 LeuPheAsnArgThrIle 492
DB 1815 ATGGTGAATTCATGGATT 1832

```

RESULT 13

```

US-08-360-673-3
; Sequence 3, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3043
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00623

```

```

; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Kluyveromyces lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 387..1862
; OTHER INFORMATION: /product= "K. lactis protease C
; OTHER INFORMATION: gene"
; OTHER INFORMATION: /gene= "kl.PRC1"
US-08-360-673-3

Alignment Scores:
Pred. No.: 1-87e-52 Length: 2503
Score: 568.50 Matches: 151
Percent Similarity: 44.65% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 184
Query Match: 19.69% Indels: 85
DB: 1 Gaps: 14

US-09-712-338-2_COPY_19_555 (1-537) x US-08-360-673-3 (1-2503)
QY 27 AlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrPhe 46
DB 570 GCTTATTCATTGAGAAATAAACCTTTGGATCCCAAACTCTCTGGCGTT-----617
QY 47 ProGlyValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePhe 66
DB 618 GATACCGTGAACAAAGTTCGGGATATTAGATACCAGGACTCAAAACACTCTTTTAT 677
QY 67 TrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGly 86
DB 678 IGGTTTTTTCAGTCTAGAAATGACCCAGAGATGACCCAGGATGATGTTTAAACGGT 737
QY 87 GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsn 106
DB 738 GGTCCTGGCTGTTCTCTCTTTCCTTGAATTTGGGACCTTCTCTCTATAGGA 797
QY 107 SerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuPhe 126
DB 798 GCTGATTTGAACCCCAATTTATAACCCCTACTCTGGAAATTCACACGCTTCTGTATTC 857
QY 127 LeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsn 146
DB 858 CTAGATCAGCTGTGGTGGTGGTCTCTCAACGGTGAC-----896
QY 147 ProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThr 166
DB 896 -----896
QY 167 IleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAlaAlaTrpGluIleLeu 186
DB 897 -----TCTAAAGTGTCTACTACAGATGACGCTGCCAAGACGTTTACATATCTTA 947
QY 187 GlnGlyPheLeuSerGlyLeuProSerIleAspSerArgValGlnSerLysAspPheSer 206
DB 948 GATTGTCTTTCGAAAGATTCCTCATTTTC-----AGAAATAACGATTTCAT 995

```


QY 62 SerHisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThr 81
 Db 1110 AAGCAITTCCTCTTTGGACTTTTGAAGTAGAACAGATCTCGAAAGGATCCGGTCAATC 1169
 QY 82 LeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGly 101
 Db 1170 CTTTGGGTGAACGGGGTCCAGGTGTCTCTCACTAACCGGGCTGTCTTGAATTAGGA 1229
 QY 102 ProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluVal 121
 Db 1230 CCTCATCTCATTTGGACCTGATTTGAACCCATCGGGAACCTTACTCTTGAACAGCAAT 1289
 QY 122 SerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrVal 141
 Db 1290 GCCACCGTGATCTCTCTTGACAGCCCTGTCAACGTTGGGTTCGTAATCC----- 1340
 QY 142 AspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 161
 Db 1341 -----GGTCTCTCA 1349
 QY 162 GlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 181
 Db 1350 GGT-----GTTCACACTGCGCGCGCTGGTAAGAT 1382
 QY 182 AlaTrpGluLeuLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGln 201
 Db 1383 GTCTATAACTCTTGGAGTGTCTTCGATCAGTCCCTGAATACGTCACAGGGCCAA 1442
 QY 202 SerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePhe 221
 Db 1443 -----GATTTCACATCGCTGGGAATCTACGCGGCATACATCCCTGTTTGGC 1496
 QY 222 AsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeu 241
 Db 1497 TCGAATAATTGCTCTCACAGGACAC----- 1523
 QY 242 AsnPheAsnSerLeuGlyIleIle-----AsnGlyIleIleAspGluAlaIleGlnAla 259
 Db 1524 AACTTCAACTTACCCGCCCTCTTGTGATCGGAATGGCTCACTGACCCATGACTCAGTAT 1583
 QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleValAlaValAsnGluThr 279
 Db 1584 AACTATTACCAACCATGGCTCTGTGTGAAGGT---GGGGAACCATCTGT----- 1631
 QY 280 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 299
 Db 1632 -----ITGCCCTCGAGGATGCTCTCTATGAGACTCTTTGGAA 1673
 QY 300 ThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu-----CysAlaGlu 317
 Db 1674 CGTTGTGGGCTGTGATCGATGCTGCAATGGTCTGCTGCTGTGCTCA 1733
 QY 318 AlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGly 337
 Db 1734 GTACCATTTATGTAAATACGCCCAATGGCTCTTACCAAGT---ACCGGCAAAAC 1790
 QY 338 ValTyrAspIleArgHisProTyrAspAsp-----ProThrProProSer 352
 Db 1791 GTTTAGGATATCAGGAAGGATTTGGAAGGTGGCAATTTGCTACCCACAGTTTACAGAT 1850
 QY 353 TyrTyrAsnLysPheLeuAlaLysSerValMetAspAlaIleGlyValAsnIleAsn 372
 Db 1851 ----ATCGACGACTACTTAACACAGGACTAGTCACAAAGAGTGTCTGCGGAGGTGAC 1907
 QY 373 TyrThrGluSer---AsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheVal 391
 Db 1908 CACTAGAAATCTGTAACTTCGATATCAACAGAAATTTCTGTTCGCGGTGATTGGATG 1967
 QY 392 TrpProAsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeu 411
 Db 1968 AAGCCT---TACCACACCCCGGTAAACAGATCTTTGAATCAACAGCTACCCATCTGGTA 2024
 QY 412 IleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAla 431

Db 2025 TATCGACGCGAIAAGATTTTCATCTGTAACTGGTGGTAATAAGGCGTGGAGGATGC 2084
 QY 432 AlaAsnTyrSerGlnAlaAlaGlnPheArgSer-----AlaGlyTyrThrProLeu 448
 Db 2085 TTGCCATGGAAGTACGAGGAAGATTTCACGCCAAAGATACGTAACCTGACCTGCTTCT 2144
 QY 449 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 468
 Db 2145 ATCACCAGCAGAGTCGCT---GGTAGCTCAATTCCTACAAGCACTTCACCTATTGAGA 2201
 QY 469 ValTyrGluAlaGlyHisGluValProTyrThrGlnProIleAlaSerLeuGlnLeuPhe 488
 Db 2202 GTCTTCATAGTGGCCACATGTTCCATTTGACGTCCTGAAACGCCCTTAAGTATGTT 2261
 QY 489 AsnArgThrIlePheGly 494
 Db 2262 AACGAATGGATCCACGGT 2279
 RESULT 15
 US-08-329-892B-32
 ; Sequence 32, Application US/08329892B
 ; Patent No. 6187579
 ; GENERAL INFORMATION:
 ; APPLICANT: Bredam, Klaus
 ; APPLICANT: Keiland-Brandt, Morten
 ; APPLICANT: Mortensen, Uffe
 ; APPLICANT: Olesen, Kjeld
 ; APPLICANT: Stennicke, Henning
 ; APPLICANT: Wagner, Fred
 ; TITLE OF INVENTION: CUSTOMIZED PROTEASE
 ; NUMBER OF SEQUENCES: 33
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
 ; STREET: 3100 No. 6187579west Center, 90 S. 7th Street
 ; CITY: Minneapolis
 ; STATE: MN
 ; COUNTRY: U.S.A.
 ; ZIP: 55402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/329,892B
 ; FILING DATE: 27-OCT-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/144,704
 ; FILING DATE: 28-OCT-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kettleberger, Denise M
 ; REGISTRATION NUMBER:
 ; REFERENCE/DOCKET NUMBER: 8648.44US01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 612/332-5300
 ; TELEFAX: 612/332-9081
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 32:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2632 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; HYPOTHEetical: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE:
 ; ORIGINAL SOURCE:
 ; FEATURE:
 ; NAME/KEY: Coding Sequence
 ; LOCATION: 696...2291

OTHER INFORMATION:
US-08-329-892B-32

Alignment Scores:

Pred. No.: 1,85e-47 Length: 2632
Score: 523.50 Matches: 147
Percent Similarity: 43.68% Conservative: 74
Best Local Similarity: 29.05% Mismatches: 200
Query Match: 18.13% Indels: 85
DB: 3 Gaps: 19

US-09-712-338-2_COPY_19_555 (1-537) x US-08-329-892B-32 (1-2632)

```
QY 15 ProLysAsnProThrGlyValIleThr-----LeuThrThrAlaAsnAsn 29
   ||||| |||
Db 939 CCAAAATCCCTCAACCAATCAACGAGACAGAGACTGGGACCTTGTGGTCAAGATGAC 998
   ||||| |||
QY 30 ValThrIleArgTyr-----LysGluProGlyAlaGluGlyVal 42
   ||||| |||
Db 999 GCAATTTGAAAGAACTATCAGCTTCGTGTCACACAGATTAAAGGACCCCTAAATTCCTGGGCATT 1058
   ||||| |||
QY 43 CysGluThrThrProGlyValIleThrSerGlyTyrValAspThrSerProGlu--- 61
   ||||| |||
Db 1059 -----GACCCAAATGTACACAGTACACGGGTTACTTGGATGTGGAAGCAGGAC 1109
   ||||| |||
QY 62 SerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThr 81
   ||||| |||
Db 1110 AAGCATTTCTCTTTGGACTTTTGAAGTGAAGACATCTCTCAAGGATCCGGTCATC 1169
   ||||| |||
QY 82 LeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGly 101
   ||||| |||
Db 1170 CTTTGGTTGAACGGGGTCCAGGTTGTTCTTCACTAACCGGGCTGTTCTTTGAATTAGGA 1229
   ||||| |||
QY 102 ProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluVal 121
   ||||| |||
Db 1230 CCCTATPCCATTCGACCTGATTTGAAACCCATCCGGAACCCCTACTCTTGGAAACAGCAAT 1289
   ||||| |||
QY 122 SerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrVal 141
   ||||| |||
Db 1290 GCACCGGTGATCTCTCTTGACAGAGCTGTCACAGCTTGGGTCTCGTATCC----- 1340
   ||||| |||
QY 142 AspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGln 161
   ||||| |||
Db 1341 -----GGGTCTCTCA 1349
   ||||| |||
QY 162 GlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 181
   ||||| |||
Db 1350 GGT-----GTTTCCACACTGTCGCCGCTGGTAAAGAT 1382
   ||||| |||
QY 182 AlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGln 201
   ||||| |||
Db 1383 GTCTATAAATCTCTGGAGTTGTTCTTCGATCAGTTCCCTGAATACGTCACAAAGGGCCAA 1442
   ||||| |||
QY 202 SerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePhe 221
   ||||| |||
Db 1443 -----GATTTCCACATCGCTGGGGAATCTACGCCGGCCATTACATCCCTGTTTTTGC 1496
   ||||| |||
QY 222 AsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeu 241
   ||||| |||
Db 1497 TCTGAATTTGTCTCACAGGACAGA----- 1523
   ||||| |||
QY 242 AsnPheAsnSerLeuGlyIleIle-----AsnGlyIleIleAspGluAlaIleGlnAla 259
   ||||| |||
Db 1524 AACTTCAACTTAACCTCCGCTTTCATCGGAATGGCCCTACTGACCCATTGACTCAGTAT 1583
   ||||| |||
QY 260 ProTyrTyrProGluPheAlaValAsnThrTyrGlyIleLysAlaValAsnGluThr 279
   ||||| |||
Db 1584 AACTATTACGAACCAATGGCCTGCGGTGAAGGT---GGCGAACCATCTGT----- 1631
   ||||| |||
QY 280 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 299
   ||||| |||
Db 1632 -----TTGCCCTCGAGGATGCTCTGCTAIGGAAGACTCTTTGGAA 1673
```

```
QY 300 ThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu-----CysAlaGlu 317
   ||||| |||
Db 1674 CGTTGTTGGGCTTGATCGAGTCTGCTATGACTCCGCAATCGCTCTGGCTGTGTCCA 1733
   ||||| |||
QY 318 AlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGly 337
   ||||| |||
Db 1734 GTTACCATTTATTGTAATACGCCCAATTCGGCTTCCTTACCACAGT---ACCGGACGANAAC 1790
   ||||| |||
QY 338 ValTyrAspIleArgHisProTyrAspAsp-----ProThrProProSer 352
   ||||| |||
Db 1791 GTTACGATATCAGGAAGGATTGTGAAGGTGGCAATTGTGTCTACCCACAGCTTACAGAT 1850
   ||||| |||
QY 353 TyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsn 372
   ||||| |||
Db 1851 ---ATCGAGACTACTTAAACACGAGCTACGTCAAAGAAAGCTGTTCGGTGGAGGTGAC 1907
   ||||| |||
QY 373 TyrThrGlnSer---AsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheVal 391
   ||||| |||
Db 1908 CACTACGAATCCCTGTACTTCGATATCAACAGAAATTCCTGTTCGGGTGATGGTGAIG 1967
   ||||| |||
QY 392 TrpProAsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeu 411
   ||||| |||
Db 1968 AAGCCT---TACCACACCGCGGTAAACAGATCTTTTGAATCAAGAGACCTACCCATCTGGTA 2024
   ||||| |||
QY 412 IleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAla 431
   ||||| |||
Db 2025 TATCGACGGCATAAAGATTTTCATCTGTAACTGGTTGGTTAATAAGCGCTGGACGGATGTC 2084
   ||||| |||
QY 432 AlaAsnTyrSerGlnAlaAlaGlnPheArgSer-----AlaGlyTyrThrProLeu 448
   ||||| |||
Db 2085 TTGCCATGGGAAGTACGACGAGAAATTTGCAAGCCAAAGTACGTACTGACTGCTTCT 2144
   ||||| |||
QY 449 LysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArg 468
   ||||| |||
Db 2145 ATCACCGACGAGGTCTGCT---GGTGAAGTCAATCTTACAAAGCACTTACCTATTGAGA 2201
   ||||| |||
QY 469 ValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuGluLeuPhe 488
   ||||| |||
Db 2202 GTCTTCAATGGTGGCCACATGGTTCCATTTCACGTCCTGAAACGCTTAAAGTATGTT 2261
   ||||| |||
QY 489 AsnArgThrIlePheGly 494
   ||||| |||
Db 2262 AAGCAATGGATCCACGGT 2279
   ||||| |||
```

Search completed: September 17, 2003, 01:46:56

Job time : 152 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:47:57 : Search time 36 seconds
(without alignments)
631.137 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555
Perfect score: 2887
Sequence: 1 LPGSTFASVGRRLPKNPTG.....HTQSSVPLPRAISMSSVGM 537

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2883	99.9	554	3	US-08-943-714-2
2	894	31.0	423	3	US-08-943-714-9
3	630	21.8	557	1	US-08-309-341-2
4	630	21.8	557	1	US-08-608-267-2
5	630	21.8	557	1	US-08-608-452-2
6	630	21.8	557	1	US-08-608-224-2
7	630	21.8	557	1	US-08-967-149-2
8	623	21.6	557	1	US-08-309-341-4
9	623	21.6	557	1	US-08-608-267-4
10	623	21.6	557	1	US-08-608-452-4
11	623	21.6	557	1	US-08-608-224-4
12	623	21.6	557	1	US-08-967-149-4
13	568.5	19.7	491	1	US-09-640-305-4
14	568.5	19.7	491	1	US-08-360-673-4
15	523.5	18.1	532	2	US-08-899-324-33
16	523.5	18.1	532	3	US-08-329-892H-33
17	513.5	17.8	421	2	US-08-607-263-4
18	361	12.5	476	4	US-08-828-488-3
19	361	12.5	476	4	US-09-299-689A-3
20	348.5	12.1	477	2	US-08-828-488-1
21	348.5	12.1	477	4	US-09-299-689A-1
22	331.5	11.5	480	2	US-08-828-488-8
23	331.5	11.5	480	4	US-09-299-689A-8
24	331.5	11.5	480	4	US-09-702-705-336
25	331.5	11.5	480	4	US-09-736-457-336
26	312.5	10.8	471	2	US-08-828-488-7
27	312.5	10.8	471	4	US-09-299-689A-7

28	266.5	9.2	523	3	US-08-943-714-11	Sequence 11, Appl
29	246.5	8.5	446	1	US-08-665-966-10	Sequence 10, Appl
30	246.5	8.5	446	3	US-09-041-780-10	Sequence 10, Appl
31	208	7.2	530	3	US-08-943-714-12	Sequence 12, Appl
32	204.5	7.1	481	3	US-08-943-714-10	Sequence 10, Appl
33	193	6.7	35	3	US-08-943-714-5	Sequence 5, Appl
34	172	6.0	179	1	US-08-665-966-8	Sequence 8, Appl
35	172	6.0	179	3	US-09-041-780-8	Sequence 8, Appl
36	169.5	5.9	351	2	US-08-828-488-5	Sequence 5, Appl
37	169.5	5.9	351	4	US-09-299-689A-5	Sequence 5, Appl
38	142	4.9	29	3	US-08-943-714-6	Sequence 6, Appl
39	127.5	4.4	771	4	US-09-634-238-229	Sequence 229, App
40	127.5	4.4	771	4	US-09-634-238-262	Sequence 262, App
41	126	4.4	385	2	US-08-387-942C-24	Sequence 24, Appl
42	126	4.4	1403	2	US-08-387-942C-3	Sequence 3, Appl
43	121	4.2	618	3	US-09-199-290-34	Sequence 34, Appl
44	120	4.2	385	2	US-08-387-942C-27	Sequence 27, Appl
45	120	4.2	872	2	US-08-387-942C-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-943-714-2
; Sequence 2, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION: 2/13/01
; APPLICANT: Blinksy, Alexander
; APPLICANT: Berk, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Gollightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Dambmann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS: 12
; ADDRESS: No. 61875780 No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,714
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENCY INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4990.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 554 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-943-714-2

Query Match 99.9%; Score 2883; DB 3; Length 554;

Best Local Similarity 100.0%; Pred. No. 1.2e-255;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFGSTPASVGRQLPKNPQKNTLTANNVTTRYKEPGAEGVCEITPGVKSYSGYDITSP 60
DB 19 LFGSTPASVGRQLPKNPQKNTLTANNVTTRYKEPGAEGVCEITPGVKSYSGYDITSP 78
QY 61 ESHTFFWFPEARHNPETAPITLWNGGPGSDSLIGLFEELGPGCHVNSTEDDYINPHSWNE 120
DB 79 ESHTFFWFPEARHNPETAPITLWNGGPGSDSLIGLFEELGPGCHVNSTEDDYINPHSWNE 138
QY 121 VSNLLFLSPLGFGSYSDTVDSINPVTGVVENSFFAGVQGRYPTIDATLIDITNLAAE 180
DB 139 VSNLLFLSPLGFGSYSDTVDSINPVTGVVENSFFAGVQGRYPTIDATLIDITNLAAE 198
QY 181 AWEILQGLSLGSLPSLDSRVQSKDFSLWTESYGHYGPAPFNHFFQNERIANGSVNGVQ 240
DB 199 AWEILQGLSLGSLPSLDSRVQSKDFSLWTESYGHYGPAPFNHFFQNERIANGSVNGVQ 258
QY 241 LFNLSGLIINGIIDEALQAPYYPPEFAVNNIYGKAYNETVYNTKMEANQMPNCQDLIST 300
DB 259 LFNLSGLIINGIIDEALQAPYYPPEFAVNNIYGKAYNETVYNTKMEANQMPNCQDLIST 318
QY 301 CKQTNRTALADYALCAEATNMCRDNVEGYPYAFAGRGVYDIRHPYDDPTPPSYNKFPAK 360
DB 319 CKQTNRTALADYALCAEATNMCRDNVEGYPYAFAGRGVYDIRHPYDDPTPPSYNKFPAK 378
QY 361 DSVMDALGVNINTQSNNDVYAFQIGDPVWPNFTEDLEILALPVRVSLIYGADAYIC 420
DB 379 DSVMDALGVNINTQSNNDVYAFQIGDPVWPNFTEDLEILALPVRVSLIYGADAYIC 438
QY 421 NWFQGGAVSLAANSQAQFRSAGYTPLVKNGVEYGETREYGNFSTRVYEAGHEVPYQ 480
DB 439 NWFQGGAVSLAANSQAQFRSAGYTPLVKNGVEYGETREYGNFSTRVYEAGHEVPYQ 498
QY 481 PIASQLQFNRTIFGWIDIAEQKKIWPSTKNGTATATHTQSSVPLPTATSMSSVGM 536
DB 499 PIASQLQFNRTIFGWIDIAEQKKIWPSTKNGTATATHTQSSVPLPTATSMSSVGM 554

RESULT 2
US-08-943-714-9
; Sequence 9, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION:
; APPLICANT: Blinkovsky, Alexander
; APPLICANT: Berk, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Gollightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Dammann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 6187578 of No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,714
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4990.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-943-714-9

Query Match 31.0%; Score 894; DB 3; Length 423;
Best Local Similarity 40.1%; Pred. No. 1.7e-73;
Matches 186; Conservative 66; Mismatches 160; Indels 52; Gaps 8;

QY 41 GVCETTPGVKSYSGYSDTVSPESHFTFFWFPEARHNPETAPITLWNGGPGSDSLIGLFEEL 100
DB 6 GICETTPGVNQSYSGYLSVGSNMNMWFWFFFEARNPQAPLAAPWPNFGPGCGSSMIGLFOEN 65
QY 101 GPGH-VNSTEDDYINPHSWNEVSNLLFLSOPLGVGFSYSDTVDSINPVTGVVENSFFAG 159
DB 66 GPCHEVNGDSTPSLNENSNWYNNYIIDQPIGVGFSYG----- 104
QY 160 VQGRYPTIDATLIDITNLAAEAEILQGLSLGSLPSLDSRVQSKDFSLWTESYGHYGP 219
DB 105 -----TDDVT---STVTAAPYVWNLQAFYAPQRYEESR---DFAITTESYGHYGP 151
QY 220 FNFHYQNERIANGSVNGVNLNLSGLIINGIIDEALQAPYYPPEFAVNNIYGKAYNET 279
DB 152 FASYIEQQAAIKAGSVTQGNVNIVALGVNNGWIDSTIQERAYIDFSYNNSYQIIDSST 211
QY 280 VYNTKMEANQMPNCQDLISTCKQTNRTALADYALCAEATNMCRDNVEGYPYAFAGRGVY 339
DB 212 RDSLLDAYN---NOCFLAQCCSSGST-----SDCINADSVCYQNTIEGPISSSDDFDY 263
QY 340 DIRHPYDDPTPPSYNKFPAKDSYMDALGVNINTQSNNDVYAFQIGDPVWPNFTEDL 399
DB 264 DIREPSNDPYPKTYSTYLSPTVVKALGARTNQECNGPYNKFASFAGDNP-RSEFLSTL 322
QY 400 EEILALPVRVSLIYGADAYICNWFQGGAVSLAANSQAQFRSAGYTPLVKNGVEYGETR 459
DB 323 SSVVQSGINVLWAGDADWICNLGNVEVANAADFPGNAQFSALDAPLAPTYNGVEKGQFK 382
QY 460 EYGNFSTRVYEAGHEVPYQPIASQLQFNRTIFGWIDIAEQKK 503
DB 383 TVDNFSLKLVYAGHEVPYQPDALQAFKQII-----QKK 418

RESULT 3
US-08-309-341-2
; Sequence 2, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5594119 of No. 5594119disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-309-341-2

Query Match          21.8%; Score 630; DB 1; Length 557;
Best Local Similarity 31.8%; Pred. No. 4.3e-49;
Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNTYIRYKPGAGVCGCTTPGVKSYGYVDTP--ESHTFFWFFEARHNPETAPITLWLN 85
Db 132 AYDLRVKKTDPGSLGI---DPGVKQYTGYYLDNENDKHLFYWFFESRNDPDPVWLWLN 188
QY 86 GPGSDSLIGLFEELGCHVNSTEDDYNPHSWNEVSNLLFLSOPLGVGFSYSTVDGSI 145
Db 189 GPGCSLTLGLFELMELGPFSSINCKTQPVNDYAWNSASVIFLDQPVNVGYYSNSA---- 244
QY 146 NPVTGVVENSFAGVQGRYPTIDATLIDTNTLAAEAAWEILQGLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDYVALLTLFFKQFP-----EYAKQDF 274
QY 206 SLWTSTGGHYGPAFFNHFYEQNERIANGSVNGVQLNFSGLIINGIIDEAIAQAPYPEF 265
Db 275 HIAGESYAGHYIPVFAEILLSHKKR-----NINLQSVLIGNGLTDGTYEYRPM 325
QY 266 AVNNTYGIKAVNEFVYVMKFANOMPNGCQDLISICKQTNKTALADYALCAEATMCRDN 325
Db 326 ACDGGYPVLDSSCCSMD--NALPR-CQSMIESCYSSA-----WVCVPASIYCNNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNLN-VTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKPEVTEAVGAEVNGVDSNCFD 436
QY 380 VYFAQGTGDFVWP--NFTEDLEELALPVRVSLIY-GDADYICNMFPGQAVSLAANSYQ 436
Db 437 INRNFLPHGDMKPYHRLVPLGLE--QIPV--LIVAGDADFICNLGNKAWTEALEWPG 491
QY 437 AAQFRSAGYTPLAV-----NGVEYGETREYGNFSFTVYVYAGHEVPPYQPIASLQFNRT 491
Db 492 QAQYASAELEDLIVDNEHTGKIGQVKSNGHETFMRLYGGGHVMPDQPESSLEFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

RESULT 4
US-08-608-267-2
; Sequence 2, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Iaver, Debbie Sue
; APPLICANT: Thompson, Cheryl Ann
; TITLE OF INVENTION: SHENI ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.

```

```

; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-608-267-2

```

```

Query Match          21.8%; Score 630; DB 1; Length 557;
Best Local Similarity 31.8%; Pred. No. 4.3e-49;
Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNTYIRYKPGAGVCGCTTPGVKSYGYVDTP--ESHTFFWFFEARHNPETAPITLWLN 85
Db 132 AYDLRVKKTDPGSLGI---DPGVKQYTGYYLDNENDKHLFYWFFESRNDPDPVWLWLN 188
QY 86 GPGSDSLIGLFEELGCHVNSTEDDYNPHSWNEVSNLLFLSOPLGVGFSYSTVDGSI 145
Db 189 GPGCSLTLGLFELMELGPFSSINCKTQPVNDYAWNSASVIFLDQPVNVGYYSNSA---- 244
QY 146 NPVTGVVENSFAGVQGRYPTIDATLIDTNTLAAEAAWEILQGLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDYVALLTLFFKQFP-----EYAKQDF 274
QY 206 SLWTSTGGHYGPAFFNHFYEQNERIANGSVNGVQLNFSGLIINGIIDEAIAQAPYPEF 265
Db 275 HIAGESYAGHYIPVFAEILLSHKKR-----NINLQSVLIGNGLTDGTYEYRPM 325
QY 266 AVNNTYGIKAVNEFVYVMKFANOMPNGCQDLISICKQTNKTALADYALCAEATMCRDN 325
Db 326 ACDGGYPVLDSSCCSMD--NALPR-CQSMIESCYSSA-----WVCVPASIYCNNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNLN-VTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKPEVTEAVGAEVNGVDSNCFD 436
QY 380 VYFAQGTGDFVWP--NFTEDLEELALPVRVSLIY-GDADYICNMFPGQAVSLAANSYQ 436
Db 437 INRNFLPHGDMKPYHRLVPLGLE--QIPV--LIVAGDADFICNLGNKAWTEALEWPG 491
QY 437 AAQFRSAGYTPLAV-----NGVEYGETREYGNFSFTVYVYAGHEVPPYQPIASLQFNRT 491
Db 492 QAQYASAELEDLIVDNEHTGKIGQVKSNGHETFMRLYGGGHVMPDQPESSLEFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

```

Db 552 LGGEW 556

RESULT 5

US-08-608-452-2

; Sequence 2, Application us/08608452

; Patent No. 5693510

; GENERAL INFORMATION:

; APPLICANT: Yaver, Debbie Sue

; APPLICANT: Thompson, Sheryl Ann

; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5693510 of No. 5693510disk of No. 5693510th America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/608,452

; FILING DATE: 28-FEB-1996

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/309,341

; FILING DATE: 20-SEP-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lowney, Karen A.

; REGISTRATION NUMBER: 31,274

; REFERENCE/DOCKET NUMBER: 4247.000-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 867 0123

; TELEFAX: 212 867 0298

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 557 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Aspergillus Niger

; US-08-608-452-2

Query Match 21.8%; Score 630; DB 1; Length 557;

Best Local Similarity 31.8%; Pred. No. 4.3e-49;

Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNVTIRKEGAGVGVVDTSP-ESHITFFWFFEARHNPETAPITLWN 85

Db 132 AYDLRVKKTDPGSLG---DPGVKQYTGVLDDNENDKHLFYWFFESRNDPNDPVVLWN 188

QY 86 GPGSDSLIGLFEELGCPCHVNSTDDYINPHSWNEVSNLLFLSQPLGVGYFSYSDTVGSI 145

Db 189 GPGGSSLTGLFMELGPSINKKIQPVYNDYAWNNSAVIFLDQPVNVGYYSNSA--- 244

QY 146 NPVTGVVNSSFAGVQGRYPTIDATLIDTTNLAFAAWELILOGFLSLPSLDSRVQSKDF 205

Db 245 -----VSDTVAACKDVIALLTLFPKQPP-----EYAKQDF 274

QY 206 SLWTSYSGHYGPAFFENHYEGNEELANGSVNGVLNLSIGLIINGIIDEALQAPYEPF 265

Db 275 HTAGESYAGHYTPVFASEILSHKK-----NINQSVLIGNGLDGTGYCYEYIRPM 325

QY 266 AVNNYTGKAVNETVYNNYKPFANOMPNGCQDLISCKQTNRPALADYALCAEATNMCRDN 325

Db 326 ACGDGGYPVLDESSCOSMD--NALPR-COSMIESCYSSEA-----WVCVPASTYCNA 377

QY 326 VEGPYAFAGRGVYVDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNLN-YTQSNND 379

Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKPEVTEAYGAENVGYDSCND 436

QY 380 VYFAFQQTGDFVWP--NFIEDLEILLALPVRVSLIY-GDADYIGNWFGGQAVSLAANYSQ 436

Db 437 INRNFLFHGDWMPKYHRLVPLLE--QIPV---LIYAGDADFICNWLGNKAWTALDWP 491

QY 437 AAOFRSAGYTPLVK-----NGVEYGETREYGNFSTRVYVYEGHVPYQPTIASLQLFNRT 491

Db 492 QAEVASAELELDVLDNEHTCKKIGQVKGSHGNFTFMRLYGGGHVMPDQPESSLEFENRW 551

QY 492 IFG-W 495

Db 552 LGGEW 556

RESULT 6

US-08-608-224-2

; Sequence 2, Application US/08608224

; Patent No. 5705376

; GENERAL INFORMATION:

; APPLICANT: Yaver, Debbie Sue

; APPLICANT: Thompson, Sheryl Ann

; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5705376 of No. 5705376disk of No. 5705376th America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/608,224

; FILING DATE: 28-FEB-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/309,341

; FILING DATE: 20-SEP-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lowney, Karen A.

; REGISTRATION NUMBER: 31,274

; REFERENCE/DOCKET NUMBER: 4247.000-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 867 0123

; TELEFAX: 212 867 0298

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 557 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Aspergillus Niger

; US-08-608-224-2

Query Match 21.8%; Score 630; DB 1; Length 557;

Best Local Similarity 31.8%; Pred. No. 4.3e-49;

Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

QY 27 ANNVTIRKEGAGVGVVDTSP-ESHITFFWFFEARHNPETAPITLWN 85

Db 132 AYDLRVKKTDPGSLG---DPGVKQYTGVLDDNENDKHLFYWFFESRNDPNDPVVLWN 188

QY 86 GPGSDSLIGLFEELGCPCHVNSTDDYINPHSWNEVSNLLFLSQPLGVGYFSYSDTVGSI 145

Db 189 GPGGSSLTGLFMELGPSINKKIQPVYNDYAWNNSAVIFLDQPVNVGYYSNSA--- 244

QY 146 NPVTGVVNSSFAGVQGRYPTIDATLIDTTNLAFAAWELILOGFLSLPSLDSRVQSKDF 205

Db 245 -----VSDTVAACKDVIALLTLFPKQPP-----EYAKQDF 274

QY 206 SLWTSYSGHYGPAFFENHYEGNEELANGSVNGVLNLSIGLIINGIIDEALQAPYEPF 265

Db 275 HTAGESYAGHYTPVFASEILSHKK-----NINQSVLIGNGLDGTGYCYEYIRPM 325

QY 266 AVNNYTGKAVNETVYNNYKPFANOMPNGCQDLISCKQTNRPALADYALCAEATNMCRDN 325

Db 326 ACGDGGYPVLDESSCOSMD--NALPR-COSMIESCYSSEA-----WVCVPASTYCNA 377

```

Db 189 GPGCCSLTGLFMEFGPSSINCKIOPVNDYAWNSASVIFLDQPVNVGYYSNSA----- 244
QY 146 NPVTGVVENSFAGVQGRYPIDATLDTTINLAEEAAWEILOGFLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDVYALLTLFFKOPF-----EYAKQDF 274
QY 206 SLWTESGGHYGPAFFNHFYEQNRIANGSVNGVQLNFNSLGIINGIIDEAIAQIAPYPEF 265
Db 275 HIAGESYAGHYIPVFASEILSHKKR-----NINLQSVLIGNGLIDGVQYQYRPM 325
QY 266 AVNNTYGIKAVNETVYNYMKFANPONGCDLISTCKQTNRTALADYALCAEATNMCRDN 325
Db 326 ACGDGGYPAVLDESSCOSMD--NALPR-COSMIESCYSSESA-----WVCVPASYCNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNNI-VTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYSDYLNKPEVIEAEGAENVGYDSCNFD 436
QY 380 VYFAFQGTGDFWVP--NFIEDLEIILALPVRSIIY-GDADYICNWFQGGQAVSLAANSQ 436
Db 437 INRNFLPHGDMKPYHRLVPLGLE--QIPV---LIYAGDADFICNWLGNKAWTEALEWPG 491
QY 437 AAFRSAGYTPLKV-----NGVEYGETREYGNFSFTRVYEGAGHEVPYQIAPSLQLENRT 491
Db 492 QAEYASAELEDLVVDNEHTGKKGIGQVKGSHGNETFMELYGCGHVMVPMQDQESSLEFFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

```

RESULT 7

```

US-08-967-149-2
; Sequence 2, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5939305o No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
US-08-967-149-2

```

```

Query Match 21.8%; Score 630; DB 2; Length 557;
Best local similarity 31.8%; Pred. No. 4.3e-49;
Matches 154; Conservative 81; Mismatches 174; Indels 76; Gaps 17;

```

```

QY 27 ANNTYIRKPEGAGVCEITPGVKSIGYVDISP-ESHITFFWFPEARNPETAPITLWLN 85
Db 132 AYDLRVKTKDPSGLI---DEGVKQYTGILDNDNKKHLEFYWFESRNDPNDPVLWLN 188
QY 86 GPGCSDSLIGLFEELGHPCHVNSTFDDYINPHSWNEVSNLLFLSOPLVGVGFYSIDTVDGS 145
Db 189 GPGCCSLTGLFMEFGPSSINCKIOPVNDYAWNSASVIFLDQPVNVGYYSNSA----- 244
QY 146 NPVTGVVENSFAGVQGRYPIDATLDTTINLAEEAAWEILOGFLSLPSLDSRVQSKDF 205
Db 245 -----VSDIVAAGKDVYALLTLFFKOPF-----EYAKQDF 274
QY 206 SLWTESGGHYGPAFFNHFYEQNRIANGSVNGVQLNFNSLGIINGIIDEAIAQIAPYPEF 265
Db 275 HIAGESYAGHYIPVFASEILSHKKR-----NINLQSVLIGNGLIDGVQYQYRPM 325
QY 266 AVNNTYGIKAVNETVYNYMKFANPONGCDLISTCKQTNRTALADYALCAEATNMCRDN 325
Db 326 ACGDGGYPAVLDESSCOSMD--NALPR-COSMIESCYSSESA-----WVCVPASYCNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVNNI-VTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYSDYLNKPEVIEAEGAENVGYDSCNFD 436
QY 380 VYFAFQGTGDFWVP--NFIEDLEIILALPVRSIIY-GDADYICNWFQGGQAVSLAANSQ 436
Db 437 INRNFLPHGDMKPYHRLVPLGLE--QIPV---LIYAGDADFICNWLGNKAWTEALEWPG 491
QY 437 AAFRSAGYTPLKV-----NGVEYGETREYGNFSFTRVYEGAGHEVPYQIAPSLQLENRT 491
Db 492 QAEYASAELEDLVVDNEHTGKKGIGQVKGSHGNETFMELYGCGHVMVPMQDQESSLEFFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

```

RESULT 8

```

US-08-309-341-4
; Sequence 4, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119o No. 5594119disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-309-341-4

Query Match 21.6%; Score 623; DB 1; Length 557;
Best Local Similarity 32.0%; Pred. No. 1.9e-48;
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

QY 27 ANNVTRYKEPGAEGCETTPGVKSYSGYVDTSP-ESHTEFFFEARHNPTAPITLWLN 85
Db 132 AYDLRVKKTDPSSLGI---DPGVKQYTGILDDNENDKHLFYWFESRNDPDPVVLWLN 188
QY 86 GGPDSLSLIGLFEELGCHVNSTFDDYINPHSWNEVSNLLFLOPLGVGFSDIVDGS 145
Db 189 GGPCCSLTGLFMELGPSSINKKIOPYNDYANNASVIFLDQPVNVGYSNSA---- 244
QY 146 NPVTGVVNSSFAGVQGRYPTIDATLIDTNLAEEAAWEILOGLFSLGLSDSRVQSKDF 205
Db 245 -----VSDTVAAGKDVYALLTLFFKQFP-----EYAKQDF 274
QY 206 SLWTESYGGHYGPAFFNHFYEQNERIANGSVNGVQLNFNSLGIINGIIDEAIAQYYPEF 265
Db 275 HIAGESYAGHYIPVFASEILSHKRR-----NINQSVLIGNGLDGLTOYEYRPM 325
QY 266 AVNNTYGIKAVNETVNYMKFANQMPNGCDLISTCKQTNRTALADYALCAEATNMCRDN 325
Db 326 ACGDG-GYPVLDLDE-GSCQAMDNALPR-CQSMIESCYSSSA-----WVCVPASIYCNA 377
QY 326 VEGPYAFAGRGVYDIRHPYDDP-----TPPSYNNKFLAKDSVMDAIGVIN-YTQSNND 379
Db 378 LLAPYQOR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKTEVIEAAGVNGYDSCNFD 436
QY 380 VYAFQOTGDFVWP--NFIEDLEILALPVRSVLIY-GDADYICNWFQGOAVSLAANSQ 436
Db 437 INRNFLHGDWMPKPYHRLVPGLE--QIPV---LIYAGDADFICNLGNKAWTEALEWPG 491
QY 437 AAQFRSAGYTPLVK-----NGVEYGETREYGNFSFTRVYAGHEVPYIQTASLQLFNR 491
Db 492 QAEYASAKLEDLVVNEHKGKKGQVKSNGHNTFMRLYGGHVMVMDQPESSLEFFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

RESULT 9
US-08-608-267-4
; Sequence 4, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
```

RESULT 10
US-08-608-452-4
; Sequence 4, Application US/08608452
; Patent No. 5693510
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5693510 of No. 5693510th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,452
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-608-452-4

Query Match 21.6%; Score 623; DB 1; Length 557;
Best Local Similarity 32.0%; Pred. NO. 1.9e-48;
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

QY 27 ANNVTIRKEPGAEGVCTTPGVKSYGVDTSP-ESHTEFFWFFFAHNPETAPITLWN 85
Db 132 AYDLRVKKTDPSSLGI--DPGVKQYGYLDNDNDKHLFYEFFERNRDPNDPVVYLWN 188
QY 86 GPGSDSLGLEELGCPCHVNSTDDYINPHSWNEVSNLLFSQPLGVGFYSYSDVDSI 145
Db 189 GPGGCSLTGLFMELGPSSINKKIOPVNDYARNASVIFLDQPVNVGYYSNSA--- 244
QY 146 NPTVGTWENSSEFAGVQGYPTIDATLIDTTLNLAEEAEILOGFLSGLPSLDSRVQSKDF 205
Db 245 -----VSDTVAAGKDVVALLTLFFKQEP-----EYAKQDF 274
QY 206 SLWTSYGGHYGAPFNHFYEQNERIANGSVGNLNSLGIINGIIDEAIQAPYYPEF 265
Db 275 HIAGESAGHYIPVASELISHKKR-----NINQSVLGNGLDGLTQYEEYRPM 325
QY 266 AVNNTYGLKAVNETVYNTWKFNQMPNGCQDLISCTKQTNRTALADYALCABATNMCRN 325
Db 326 AGCGDQ-GYPAVLDE-GSCQAMDNALPR-CQSMIESCYSSESA-----WVCVPASICYNA 377
QY 326 VEGPYVAFAGRGVDIRPYDDP-----TPPSYNNKFLAKUSVMDAIGNVIN-YTQSNND 379
Db 378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKTEVIEAVGAENVGYDSCNFD 436

QY 380 VYAFQQTGDFWVP--NFIEDLEBETLALPVRVSLIY-GDADYICNWFQGAQVSLAANYSQ 436
Db 437 INRNFLFHGDWMPYHRLVPGILLE--QIPV---LIYAGDAOFICNWLGNKAWTEALEWPG 491
QY 437 AAQFRSAGYTPKV-----NGVEYGETREYGNFSEFTRYVEAGHEVPPYQPIASLOLQFNRT 491
Db 492 QAEYASAKLEDLVVNEHKKIGQVKSNGNFTMRDLYGGHVMVMDQDPSSLEFFNRW 551
QY 492 IFG-W 495
Db 552 LGGEW 556

RESULT 11
US-08-608-224-4
; Sequence 4, Application US/08608224
; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5705376 of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
; US-08-608-224-4

Query Match 21.6%; Score 623; DB 1; Length 557;
Best Local Similarity 32.0%; Pred. NO. 1.9e-48;
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

QY 27 ANNVTIRKEPGAEGVCTTPGVKSYGVDTSP-ESHTEFFWFFFAHNPETAPITLWN 85
Db 132 AYDLRVKKTDPSSLGI--DPGVKQYGYLDNDNDKHLFYEFFERNRDPNDPVVYLWN 188
QY 86 GPGSDSLGLEELGCPCHVNSTDDYINPHSWNEVSNLLFSQPLGVGFYSYSDVDSI 145
Db 189 GPGGCSLTGLFMELGPSSINKKIOPVNDYARNASVIFLDQPVNVGYYSNSA--- 244
QY 146 NPTVGTWENSSEFAGVQGYPTIDATLIDTTLNLAEEAEILOGFLSGLPSLDSRVQSKDF 205

ORIGINAL SOURCE:
ORGANISM: Aspergillus Niger
US-08-967-149-4

Query Match 21.6%; Score 623; DB 2; Length 557;
Best Local Similarity 32.0%; Pred. No. 1.9e-48;
Matches 155; Conservative 79; Mismatches 175; Indels 76; Gaps 18;

27 ANNVTIRYKPGAGVCEETPGVKSYGYYVDTSP-ESHTFFWFEARHNPETAPILWLN 85
132 AYDLRVKTKDPSSLGI--DPGVKQYTGYYLDDNENKHLFWFEFESRNDPENVVWLN 188

86 GPGCSLSLGLFEELGPGCHVNSTFDDYINPHSNEYSNLLFLSOPLGFGVSYSDTVGSI 145
189 GPGCSLSLGLFEELGPGCHVNSTFDDYINPHSNEYSNLLFLSOPLGFGVSYSDTVGSI 244

146 NPVTGVVENSFAGVQGRYPTIDATILDTFTNLAEEAEWEILOGLSLGSLDSRVOSKOF 205
245 -----VSDTVAAGKDVYALLTLFFKQFP----EYAKQDF 274

206 SLMTESYGGHYGPAFFNHFYEQNERANGSVNGVQLNFNSLGIINGLIIDEALQAPYYPF 265
275 HIAGESYAGHYIPVFASEILSHKKR-----NINLQSVLIGNGLTDLGTOYEYRPM 325

266 AVNNTYGLKAVNETVYNYMKFANOMPNGCODLSTCKOTNKTALADYALCAEATNMCRDN 325
326 ACGDG-GYPAVLDE-GSCQAMDNALPR-COSMIESCYSSESA-----WVCVPASLYCNA 377

326 VEGFYAFAGRGVYDIRHPYDDP-----TPSYNKFELAKDSVMDAIGVNNIN-YTQSNND 379
378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKTEVIEAVGAEVNGYDSCNFD 436

380 VYAFQOTGDFWVP--NFIEDLEILALPVRVSLIY-GDADYICNWFQGAQVLAANYSQ 436
437 INRNFLHGDWMPKPYHRLVPGLE--QIPV--LIYAGDADFICNLGNKAWTEALEWPG 491

437 AAQFRSAGYTPLKV-----NGVEYGETREYGNFSTFVYAGHEVPYQPIASLQLFNRT 491
492 QAEYASAKLEDLVVNEHKGKKGQVKSNGNFTMRLYGGGHMVPMDQPESSLEFFNRW 551

492 IFG-W 495
552 LGGEW 556

RESULT 13
US-09-640-305-4
Sequence 4, Application US/09640305
Patent No. RE37447
GENERAL INFORMATION:
APPLICANT: Fleer, Reinhard
Fournier, Alain
Yeh, Patrice
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR PREPARATION AND USE
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Rd. 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19002
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/640,305
FILING DATE: 16-Aug-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/360,673

245 -----VSDTVAAGKDVYALLTLFFKQFP----EYAKQDF 274

206 SLMTESYGGHYGPAFFNHFYEQNERANGSVNGVQLNFNSLGIINGLIIDEALQAPYYPF 265
275 HIAGESYAGHYIPVFASEILSHKKR-----NINLQSVLIGNGLTDLGTOYEYRPM 325

266 AVNNTYGLKAVNETVYNYMKFANOMPNGCODLSTCKOTNKTALADYALCAEATNMCRDN 325
326 ACGDG-GYPAVLDE-GSCQAMDNALPR-COSMIESCYSSESA-----WVCVPASLYCNA 377

326 VEGFYAFAGRGVYDIRHPYDDP-----TPSYNKFELAKDSVMDAIGVNNIN-YTQSNND 379
378 LLAPYQR-TGQNVYDVRGKCEDSSNLCYSAMGYVSDYLNKTEVIEAVGAEVNGYDSCNFD 436

380 VYAFQOTGDFWVP--NFIEDLEILALPVRVSLIY-GDADYICNWFQGAQVLAANYSQ 436
437 INRNFLHGDWMPKPYHRLVPGLE--QIPV--LIYAGDADFICNLGNKAWTEALEWPG 491

437 AAQFRSAGYTPLKV-----NGVEYGETREYGNFSTFVYAGHEVPYQPIASLQLFNRT 491
492 QAEYASAKLEDLVVNEHKGKKGQVKSNGNFTMRLYGGGHMVPMDQPESSLEFFNRW 551

492 IFG-W 495
552 LGGEW 556

RESULT 12
US-08-967-149-4
Sequence 4, Application US/08967149
Patent No. 5939305
GENERAL INFORMATION:
APPLICANT: Yaver, Debbie Sue
Thompson, Sheryl Ann
TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59393050 No. 59393050disk of No. 59393050th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,149
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,452
FILING DATE: 28-FEB-1996
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 557 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein


```

; FILING DATE: 06-FEB-1995
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Smith, Julie K.
;   REGISTRATION NUMBER: 38,619
;   REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (610)454-3839
;   TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 491 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-640-305-4

Query Match          19.7%; Score 568.5; DB 1; Length 491;
Best Local Similarity 31.1%; Pred. No. 1.5e-43;
Matches 151; Conservative 66; Mismatches 184; Indels 85; Gaps 14;

QY 27 ANVTIRYKPPGAGVCEITPGVKSYSYVDTSPESTFFWFPEARHNPTAPITLWNG 86
Db 62 AYSLRIKPLDPKSLGV---DTVKQWSGYLDYQDSKHHFFWFESRNDPNDPVILWNG 117
QY 87 GPCSDSLIGLFEELGCHVNSTFDDYINPHSWNEVSNLLFSLQPLGVGFYSYDVTGDSIN 146
Db 118 GPCSSFVGLFFELGPGSSIGADLKPIYNPYSWNSNASVIFLDQPVGVGFSGYD----- 170
QY 147 PVTGVVENSFAGVQGRYPTIDATLIDTNLAEEAAWEILQGLSLGLPSLDSRVQSKDFS 206
Db 171 -----SKVSTDDAAKDVYIFLDLFFERFPHL-----RNNDHF 203
QY 207 LWTESYGGHYGPAFFNHFEQNERIANGSVNGVQLNFNSLGIINGIDEAIOAPYIPEFA 266
Db 204 ISGESYAGHYLPKIAH-----EIAVVAEDSSFNLSVVLGNGFTDPLTQYIYEPMA 256
QY 267 VNNTYGIKAVNETVYNYMKFANOMPNGCOD-----LISTCKOTNTALADYALCA 316
Db 257 CGEG-GYPVAVLE-----PEDCLDMNENLPLCLSLVDRCYKSH-----SVFSCV 298
QY 317 EATNMCRDVNEGPYAFAGGVYDIR---HPYDDP---TPPSYNNKFLAKDSVMDAIGV 369
Db 299 LADRYCEQQTIG-VYEKSGRNPYDIRSKCEAEDDSGACYQEETIYSDYLNQEEVQRAIGT 357
QY 370 NINYTOS-NDVYVAFQQTGDFVWPNFIEDLEILALPVRSVLIYGDADYICNWFQGVAV 428
Db 358 DVSSFGCCSDVGIGFAFTGDGSP-FHQYVVAELLDDQDINVLVYAGDKYICNWLGNLAW 416
QY 429 SLAANYSQAAQFSAAGYTPKLVNGVE--YGETREYGNFSETRVYAGHEVPYIYQPIASLQ 486
Db 417 TEKLEWRYNEEYKKQVLRWTWKSEETDETIGETKSYGLYLIYRIYDAGHVMVPHDPENSILQ 476
QY 487 LENRTI 492
Db 477 MYNSWI 482

RESULT 14
US-08-360-673-4
; Sequence 4, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3c43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 491 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-360-673-4

```

```

Query Match          19.7%; Score 568.5; DB 1; Length 491;
Best Local Similarity 31.1%; Pred. No. 1.5e-43;
Matches 151; Conservative 66; Mismatches 184; Indels 85; Gaps 14;

QY 27 ANVTIRYKPPGAGVCEITPGVKSYSYVDTSPESTFFWFPEARHNPTAPITLWNG 86
Db 62 AYSLRIKPLDPKSLGV---DTVKQWSGYLDYQDSKHHFFWFESRNDPNDPVILWNG 117
QY 87 GPCSDSLIGLFEELGCHVNSTFDDYINPHSWNEVSNLLFSLQPLGVGFYSYDVTGDSIN 146
Db 118 GPCSSFVGLFFELGPGSSIGADLKPIYNPYSWNSNASVIFLDQPVGVGFSGYD----- 170
QY 147 PVTGVVENSFAGVQGRYPTIDATLIDTNLAEEAAWEILQGLSLGLPSLDSRVQSKDFS 206
Db 171 -----SKVSTDDAAKDVYIFLDLFFERFPHL-----RNNDHF 203
QY 207 LWTESYGGHYGPAFFNHFEQNERIANGSVNGVQLNFNSLGIINGIDEAIOAPYIPEFA 266
Db 204 ISGESYAGHYLPKIAH-----EIAVVAEDSSFNLSVVLGNGFTDPLTQYIYEPMA 256
QY 267 VNNTYGIKAVNETVYNYMKFANOMPNGCOD-----LISTCKOTNTALADYALCA 316
Db 257 CGEG-GYPVAVLE-----PEDCLDMNENLPLCLSLVDRCYKSH-----SVFSCV 298
QY 317 EATNMCRDVNEGPYAFAGGVYDIR---HPYDDP---TPPSYNNKFLAKDSVMDAIGV 369
Db 299 LADRYCEQQTIG-VYEKSGRNPYDIRSKCEAEDDSGACYQEETIYSDYLNQEEVQRAIGT 357
QY 370 NINYTOS-NDVYVAFQQTGDFVWPNFIEDLEILALPVRSVLIYGDADYICNWFQGVAV 428
Db 358 DVSSFGCCSDVGIGFAFTGDGSP-FHQYVVAELLDDQDINVLVYAGDKYICNWLGNLAW 416
QY 429 SLAANYSQAAQFSAAGYTPKLVNGVE--YGETREYGNFSETRVYAGHEVPYIYQPIASLQ 486
Db 417 TEKLEWRYNEEYKKQVLRWTWKSEETDETIGETKSYGLYLIYRIYDAGHVMVPHDPENSILQ 476

```

Search completed: September 16, 2003, 22:53:19
Job time : 38 secs

Query Match	18.1%	Score 523.5	DB 2:	Length 532;
Best Local Similarity	29.1%	Pred. No. 2.3e-39;		
Matches 147;	Conservative 74;	Mismatches 200;	Indels 85;	Gaps 19;
15	PKNPTGVKT-----UTANNVTIR-----KFGAECVCTTFCVKSYGVVDISPE- 61			
DB	82	PKFPEAIKTKKDWEVWKNDALENQLRVNKKIDPKILGI---DPNVTQITGYLQVED 138		
16	SHTFWFPEARHNPETAPITLWNLGPGSDSLIGLFEELGPGCHVNSTFDDYNPHSWNEV 121			
QY	139	KHFFFTWTFESRNDPAKDPVILWNLGPGGSSITGTFEELGPGSSIGDPLKIPGNPYSWNSN 198		
DB				
17	SNLLFSLQPLGFGFYSYDVSIVGDSINPVTGVWVNSSFAGVQGRYPTIDATLIDTINLAAE 181			

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 16, 2003, 18:42:04 ; Search time 28 Seconds
(without alignments)
4859.704 Million cell updates/sec
Title: US-09-712-338-1_COPY_55_1662
Perfect score: 2904
Sequence: 1 cttccaggaagtacaccggc.....gcattgccagtgttgatg 1608

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DBV=xlp
-Q=/cn2_1/USPTO.spool.p/US09712338/runat_16092003_144326_14403/app_query.fasta_1.1799
-DB=Issued Patents.AA -QFMT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAIRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338.@CN_1_1_31.@runat_16092003_144326_14403 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.AA.*
1: /cn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cn2_6/ptodata/2/iaa/6C.COMB.pep.*
6: /cn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	2878	99.1	554	3	US-08-943-714-2
2	894	30.8	423	3	US-08-943-714-9
3	630	21.7	557	1	US-08-309-341-2
4	630	21.7	557	1	US-08-309-341-2
5	630	21.7	557	1	US-08-608-267-2
6	630	21.7	557	1	US-08-608-267-2
7	630	21.7	557	1	US-08-608-224-2
8	623	21.5	557	1	US-08-967-149-2
9	623	21.5	557	1	US-08-309-341-4
10	623	21.5	557	1	US-08-608-267-4
11	623	21.5	557	1	US-08-608-267-4
12	623	21.5	557	2	US-08-967-149-4

SUMMARIES

13	568.5	19.6	491	1	US-09-640-305-4	Sequence 4, Appli
14	568.5	19.6	491	1	US-08-360-673-4	Sequence 4, Appli
15	523.5	18.0	532	3	US-08-899-324-33	Sequence 33, Appl
16	523.5	18.0	532	3	US-08-329-892B-33	Sequence 33, Appl
17	513.5	17.4	421	2	US-08-807-263-4	Sequence 4, Appli
18	361	12.4	476	2	US-08-828-488-3	Sequence 3, Appli
19	361	12.4	476	4	US-09-299-689A-3	Sequence 3, Appli
20	348.5	12.0	477	2	US-08-828-488-1	Sequence 1, Appli
21	348.5	12.0	477	4	US-09-299-689A-1	Sequence 1, Appli
22	332.5	11.4	480	2	US-08-828-488-8	Sequence 8, Appli
23	332.5	11.4	480	4	US-09-299-689A-8	Sequence 8, Appli
24	332.5	11.4	480	4	US-09-702-705-336	Sequence 336, App
25	332.5	11.4	480	4	US-09-736-457-336	Sequence 336, App
26	312.5	10.8	471	2	US-08-828-488-7	Sequence 7, Appli
27	312.5	10.8	471	4	US-09-299-689A-7	Sequence 7, Appli
28	266.5	9.2	523	3	US-08-943-714-11	Sequence 11, Appli
29	246.5	8.5	446	1	US-08-665-966-10	Sequence 10, Appl
30	246.5	8.5	446	3	US-09-041-780-10	Sequence 10, Appl
31	208	7.2	530	3	US-08-943-714-12	Sequence 12, Appl
32	204.5	7.0	481	3	US-08-943-714-10	Sequence 10, Appl
33	193	6.6	35	3	US-08-943-714-5	Sequence 5, Appli
34	172	5.9	179	1	US-08-665-966-8	Sequence 8, Appli
35	172	5.9	179	3	US-09-041-780-8	Sequence 8, Appli
36	169.5	5.8	351	2	US-08-828-488-5	Sequence 5, Appli
37	169.5	5.8	351	4	US-09-299-689A-5	Sequence 5, Appli
38	163.5	5.6	878	4	US-09-536-706B-2	Sequence 2, Appli
39	161	5.5	907	3	US-08-783-774-2	Sequence 2, Appli
40	161	5.5	907	4	US-09-328-599A-1	Sequence 1, Appli
41	161	5.5	907	5	PCT-US95-04611A-19	Sequence 19, Appl
42	156	5.4	1048	4	US-09-171-699-10	Sequence 10, Appl
43	154	5.3	1911	4	US-09-854-856-64	Sequence 64, Appl
44	154	5.3	1939	4	US-09-854-856-48	Sequence 48, Appl
45	154	5.3	1971	4	US-09-854-856-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-943-714-2
; Sequence 2, Application US/08943714
; Patent No. 6187578
; GENERAL INFORMATION:
; APPLICANT: Blinkovsky, Alexander
; APPLICANT: Berk, Randy
; APPLICANT: Rey, Michael
; APPLICANT: Golightly, Elizabeth
; APPLICANT: Klotz, Alan
; APPLICANT: Mathisen, Thomas Erik
; APPLICANT: Danbmann, Claus
; TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 6187578o No. 6187578disk of No. 6187578th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,714
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4990.200-US
; TELECOMMUNICATION INFORMATION:

102e-1000000000

102e-1000000000

102e

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 554 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-943-714-2

Alignment Scores:
Pred. No.: 1,296-268 Length: 554
Score: 2878.00 Matches: 536
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.10% Indels: 0
DB: 3 Gaps: 0

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-943-714-2 (1-554)

QY 1 CTTCCAGGAGTACACCGCGCTCGGTGCTAGAACACAGTACCCCAAGAACCCCGGG 60
DB 19 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 38
QY 61 GTCAGACCTTACACCCCAACATGTCACCATCCGGTACAGGAACCCCGGGCAGAG 120
DB 39 ValLysThrLeuThrThraAlaAsnValThrileArgTyrLysGluProGlyAlaGlu 58
QY 121 GGCGTCGGAGACTACCCCGGTGTCAAATCCTACTCTGGATATGTCGACACCTCTCC 180
DB 59 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 78
QY 181 GAGTCCCATACCTTCTCTGGTCTTCGAGCCAGACATACCCAGAACTGCACCTATC 240
DB 79 GluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIle 98
QY 241 ACATTGGTGGTGAATCGTGGCCCTGGAAGGATCTTTGATCGTCTCTCGAAGATTG 300
DB 99 ThrLeuTrpLeuAsnGlyProGlySerAspSerLeuileGlyLeuPheGluGluLeu 118
QY 301 GGCGCTGGCATCTCAATTCGACTTTGATGACTATCATCAACCTCCTCGTGGAGCAG 360
DB 119 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu 138
QY 361 GTCTCCCAATTTACTATTCCTGTCCTCCAGCCCATTTGGAGTCGGCTTTTCATATAGTATCAG 420
DB 139 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 158
QY 421 GTTCATGGTCCATTACCTGTAACCTGGGCTGCGAAATTCGAGCTTTCGAGGAGTT 480
DB 159 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 178
QY 481 CAGGCGCGTACCAACCATTTGATGCGACTCTGATGATACCTACCAATCTTGGCGCAGAG 540
DB 179 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu 198
QY 541 GCCGCTGGAGATCTCGAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGGIG 600
DB 199 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 218
QY 601 CAGTCTAAGGACTTCACTGATGACGAGGAGCTATGAGGCGCACTATGTCCTGCATTC 660
DB 219 GlnSerLysAspPheSerLeuThrThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 238
QY 661 TTCATCATTTTACGACGAGATGAGAGAATGCCAACGGTACTGTATTAATGTTTCAG 720
DB 239 PheAsnHisPheTyrGluGluAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 258
QY 721 CTTAATTTCACTCTCTGGGAATTTAATTAACGGCATTCAGCAGGCGGATCCAGCCCT 780

DB 259 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro 278
QY 781 TACTACCCCTGAATTCGCTGTGAACATACCTACGGTATCAAGGCTGTCAACGAGACGCTC 840
DB 279 TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleYsAlaValAsnGluThrVal 298
QY 841 TACAACATACATGAAGTTTCCCAACAAATGCCAAATGGTATGGTGGAGATTGATTCCACC 900
DB 299 TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 318
QY 901 TGCARACACACAAACCGCAGCCGCTAGCTAGCTAGCCCTCTGCGCCGGAAGCACCACAC 960
DB 319 CysLysGlnThrAsnArgThrAlaLeuAlaAspThrAlaLeuCysAlaGluAlaThrAsn 338
QY 961 ATGTGACGGGACAATGTGAGGGCCATACCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 339 MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 358
QY 1021 ATTCGCGATCCATATGATGACCGGCTCGGCAAGTATTTACAAATTTCTGGCAAG 1080
DB 359 IleArgHisProTyrAspAspProThrProSerTyrTyrAsnLysPheLeuAlaLys 378
QY 1081 GACTCTGTCATGACGCTATCGGCTCAACATCACTACACCCAGCTCCATATATGAGTC 1140
DB 379 AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 398
QY 1141 TACTAGCTTTCACGCAACACAGCGACTTTGCTGCGCCCACTCATCTGCGAGACCTCGAG 1200
DB 399 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 418
QY 1201 GAGATCCTGCTCCTCCCGCTGCGTCTCCTCATCTATGCGACGCGGATTCACATCTGC 1260
DB 419 GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 438
QY 1261 AACTGTTGCTGCGCTAGCGCTTTCCTCGTGGAACTACTCCCAAGCGCGCCAGTTC 1320
DB 439 AsnTrpPheGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 458
QY 1321 CGAAGCGCAGGTCACGCGCCCTGAAAGTCAAGCGCTCGAGTATGGGGAACCTCGCGAG 1380
DB 459 ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 478
QY 1381 TATGGTAATTTCTCTTCACTCGGCTATGAGCGAGCGGCTAGGCTCCCATCTACTACAG 1440
DB 479 TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 498
QY 1441 CCCATCGCTCCCTCGCAATTTTAAACGGACTATCTTCTGGTGGGATATCGCAGAGGC 1500
DB 499 ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 518
QY 1501 CAGAAGAAGATCTCGCCGAGCTACAGACGAGTGGAGCGGCTAGCTAGCTAGCTAGCTAG 1560
DB 519 GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln 538
QY 1561 TCGTCCGTCGCGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT 1608
DB 539 SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet 554

RESULT 2
US-08-943-714-9
Sequence 9, Application US/08943714
Patent No. 6187578
GENERAL INFORMATION:
APPLICANT: Blinkovsky, Alexander
APPLICANT: Berk, Randy
APPLICANT: Rey, Michael
APPLICANT: Golightly, Elizabeth
APPLICANT: Klotz, Alan
APPLICANT: Mathisen, Thomas Erik
APPLICANT: Dammann, Claus
TITLE OF INVENTION: Carboxypeptidases And Nucleic Acids
TITLE OF INVENTION: Encoding Same
NUMBER OF SEQUENCES: 12


```

US-08-608-267-2
; Sequence 2, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
US-08-608-267-2

Alignment Scores:
Pred. No.:          6,13e-52
Score:             630.00
Percent Similarity: 48.45%
Best Local Similarity: 31.75%
Query Match:       21.69%
DB:                1

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-267-2 (1-557)
QY      79  GCACCAATGTCACCATCCGGTACAAGAACCGGGGCAGAGGCGTGTCGAGACTACC 138
Db      ||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
        132  AlalyrAspLeuArgValLysThrAspProglySerLeuGlyIle-----Asp 148
QY      139  CGGGGTGCTAAATCCTACTTCGATATGTCGACACCTCTGCC---GAGTCCCATACTGTGGTTGAAT 195
Db      |||||:||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
        149  ProGLyVallysGlnIyrThrGlyrLeuAspAsnGluAsnAspLysHisLeuPhe 168
QY      196  TTCTGGTTCTTTCGAAGCCACACATAACCCAGAACCTGCACCTATCACATCTGTGGTTGAAT 255
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
        169  TyrTrpPheGluSerArgAsnAspProgluAsnAspProvalValLeuTrpLeuAsn 188
QY      256  GGTTGCCCTGGAGAGCATICTTTGATCGGTCTCTTCGAAGAGTTGGCCCTTGGCCATGTC 315
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
        189  GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY      316  AATTTCGACITTTGATGACTACATCACCCCTCACTCGTGGGAACGAGGTTCTCCAATTACIA 375
Db      ||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

```

QY 1414 GCAGGCCATGAGTCCCTACTACAGCCATCGCTCCCTGCAATGTTTACCGGACT 1473
 Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTirp 551
 QY 1474 ATCTCGGT---TGG 1485
 Db 552 LeuGlyGlyGluTirp 556

RESULT 5
 US-08-608-452-2
 ; Sequence 2, Application US/08608452
 ; Patent No. 5693510
 ; GENERAL INFORMATION:
 ; APPLICANT: Yaver, Debbie Sue
 ; APPLICANT: Thompson, Sheryl Ann
 ; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 56935100 No. 5693510disk of No. 5693510th America, Inc.
 ; STREET: 405 Lexington Avenue, Suite 6400
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/608,452
 ; FILING DATE: 28-FEB-1996
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/309,341
 ; FILING DATE: 20-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lowney, Karen A.
 ; REGISTRATION NUMBER: 31,274
 ; REFERENCE/DOCKET NUMBER: 4247.000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 867 0123
 ; TELEFAX: 212 867 0298
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 557 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Aspergillus Niger
 ; US-08-608-452-2

Alignment Scores:
 Pred. No.: 6,13e-52
 Score: 630.00
 Percent Similarity: 48.45%
 Best Local Similarity: 31.75%
 Query Match: 21.69%
 DB: 1

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-452-2 (1-557)
 QY 79 GCAACAAATGTCACATCGGTACAGAACCCGGGCGAGAGCGGTCTGGGACTACC 138
 Db 132 AlaTyrAspLeuArgValLysThrSerProGlySerLeuGlyIle-----Asp 148
 QY 139 CCGGTGTCAATCTACTCTGGATATGTCGACACCTCTCCG---GAGTCCCATACCTTC 195
 Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 168

QY 196 TTCTGGTTCTTGAAGCCAGACATACCCAGAAACTGCACCTATCACAATTCGTGTTGAAT 255
 Db 169 TyrTirpPheGluSerArgAsnAspProGluAsnAspProValValLeuTirpLeuAsn 188
 QY 256 GGTGGCCCTGGAACCGATTCTTTCATCGGTCTCTCGAAGAGTTGGCCCTTGCATCTC 315
 Db 189 GlyGlyProGlyCysSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
 QY 316 AATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGAAACGAGGTCTCCAAATTACTA 375
 Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTirpAsnSerAsnAlaSerValIle 228
 QY 376 TTCTGTGTCGCCAGCCATTGGGAGTGGCTTTTCATATAGTACAGGTGATGGTCCCAT 435
 Db 229 PheLeuAspGlnProValAsnValClyTyrSerTyrSerAsnSerAla----- 244
 QY 436 AACCTGTAACTGGGGTCTCGAAATTCGAGCTTTCAGGAGTTTCAGGGCGGTACCCA 495
 Db 244 ----- 244
 QY 496 ACCATTGATGCCACTCTGATCGATACIACCAATCTTGGCGAGAGCGGCTTGGAGATC 555
 Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
 QY 556 CTGCAAGGATTCCTTAGTGGACTACCTAGCTTGGACTCTAGGTGCTAGCTTCAAGACTTC 615
 Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274
 QY 616 ACTCTATGAGCGGAGAGCTATGAGGGCTACTATGGTCTGCATCTTCAATCATTTTAC 675
 Db 275 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294
 QY 676 GAGCAGAAAGAGAGATTGCCAAGCTAGTGTAAATGTGTTCAGCTTAATTTCAACTCT 735
 Db 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
 QY 736 CTGGGAATTATTAAACGGCATATPCAGAGGCGCATCCAGGCCCTTACTACCTGATTC 795
 Db 306 ValLeuIleGlyAsnGlyLeuThrAspGlyTyrThrGlnTyrGluTyrTyrArgProMet 325
 QY 796 GCTGTGAACAATACCTACGTATCAAGGTCTCAAGGAGACCTCTACAACTACGAAG 855
 Db 326 AlaCysGlyAspGlyGlyTyrProAlaValLeuAspGluSerSerCysGlnSerMetAsp 345
 QY 856 TTTCGCCAACCAATGCCAATGGTTCGACGAGTTGATTTCCACCTGCAACAGACAAAC 915
 Db 346 -----AsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerGlu 362
 QY 916 CGCACCGCATTAGCTAGCTACGCCCTCTGCGCCGAGACCCACACATCTCAGGAGCAAT 975
 Db 363 SerAla-----TipValCysValProAlaSerIleTyrCysAsnAla 377
 QY 976 GTTCAGGGGCCATACACTAGCCCTTTCGTGTGTGTGTATGATATTCGSCATCCATAT 1035
 Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396
 QY 1036 GATCACCG-----ACTCCGCCAAGTTATTACAAATTTCTGGCAAG 1080
 Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
 QY 1081 GACTCTGTATGAGCGGTATCGGGCTCAACATCAAC---IACACCCAGTCCCAATAATGAC 1137
 Db 417 ProGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436
 QY 1138 GTCTACTACGCTTCCACCAACAGGCGACTTTGTCTGGCC-----AATCTCATCGAA 1191
 Db 437 IleAsnArgAsnPheLeuPheHisGlyAspTyrMetLysProTyrHisArgLeuValPro 456
 QY 1192 GACCTCGAGGAGATCCTTCTCTCCCTCCCGGTGTCTCCCTCATCTAT---GGCAGCCG 1248
 Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
 QY 1249 GATTACATCTGCAACTGGTTTCGGCGGTTCAGGGCGGTTTCCTCTCGCTCGCAACTACTCCCA 1308


```

Db      472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
QY      1309 GCGCCCGAGTCCGAAAGCGAGGTACAGCCCTCGAAAGTC-----AAC 1353
Db      492 GluAlaGluTrpAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr 511
QY      1354 GCGCTCGAGTATGGGAAACTCGGAGTATGGTAATTCCTCTCACTCGCGTCTATGAG 1413
Db      512 GlnLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTrpGly 531
QY      1414 GCGAGGCATGAAGTCCCATCTACTACCGCCATCCCTCCCTCGCAATTTGTTTAAACCGGACT 1473
Db      532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTyr 551
QY      1474 ATCTTCGGT---TGG 1485
Db      552 LeuGlyGlyGluTrp 556

```

RESULT 6

```

US-08-608-224-2
; Sequence 2, Application US/08608224
; Patent No. 5705376
; GENERAL INFORMATION:
; APPLICANT: Iaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5705376 No. 5705376disk of No. 5705376th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,224
; FILING DATE: 28-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
US-08-608-224-2

```

```

Alignment Scores:
Pred. No.: 6,13e-52
Score: 630.00
Length: 557
Matches: 154
Percent Similarity: 48.45%
Conservative: 81
Best Local Similarity: 31.75%
Mismatch: 174
Query Match: 21.69%
Indels: 76
DB: 1
Gaps: 17

```

```

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-224-2 (1-557)
QY      79 GCAACAATGTCACCATCCGGTACAAAGAACCGGGGAGAGGGGCTCTGCGAGACTACC 138
Db      132 AlaTyrAspLeuArgValLysLysThrAspProGlySerLeuGlyIle-----Asp 148
QY      139 CCGGGTGTCAAACTCCTACCTCTGATATGTCACACCTCTCC---GAGTCCCATACCTTC 195
Db      149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAsnGluAsnAspLysHisLeuPhe 168
QY      196 TTCTGGTCTTCGAAGCAGACATAACCCAGAACTGCACCTATACACATCTGTTGAT 255
Db      169 TyrTrpPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
QY      256 GGTGGCCCTCGAAGCGATTCTTTGATCGGTCTCTTCGAAGAGTGGGCGCTTCGCATGTC 315
Db      189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY      316 AATTGCGACTTTTGATGACTACATCAACCTCTACCTCGTGGACAGAGGTCTCCAAATTACTA 375
Db      209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
QY      376 TTCTGTCCCGAGCCATTGGGAGTCGGCTTTTCATATAGTAGTACATCGTTGATGGTCCATT 435
Db      229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla----- 244
QY      436 AACCTCTAACTGGGCTCGTGGAAATTCGAGCTTTCGAGAGTTCAGGGCGCGGTACCCA 495
Db      244 ----- 244
QY      496 ACCATTGATGCCACTCTGATCGATACACCAATCTTCGCGAGAGGCGGTGGAGATC 555
Db      245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
QY      556 CTGCAAGGATTCTCTAGTGGACTACCPAGCTTGGACTCTAGGCTGAGGTGCGACTCTAAGGACTTC 615
Db      259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274
QY      616 AGTCTATGGACGAGAGCTATGGAGGCGACTATGGCTCTGCTCATCTTCAATCATTTTAC 675
Db      275 HisIleAlaGlyGluSerTyrAlaGlyHisIleProValPheAlaSerGluIleLeu 294
QY      676 GACGAGAATGAGAGATTCGCCAAGGTAGTGTAAATGGTTCAGCTTAATTTCAACTCT 735
Db      295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
QY      736 CTGGGAATTATTAAACGGCATCATCGAGGCGGATCCAGGCGCTTACTTACTTACCCTGAATTC 795
Db      306 ValLeuIleGlyAsnGlyLeuThrAspGlyTyrThrGlnTyrGluTyrTyrArgPromet 325
QY      796 GCTGTGAACAATATACCGTATCAAGGCTGTCAAGGAGACCGCTCTACAACACTACATCAAG 855
Db      326 AlaCysGlyAspGlyGlyTyrProAlaValLeuAspLysSerSerCysGlnSerMetAsp 345
QY      856 TTTCGCCAACCAATGCCAATGTTGCCAGGATTTGATTTCCACCTGCAACAGACAAC 915
Db      346 -----AsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerGlu 362
QY      916 CGCAGCGCATTAGCTGACTAGCGCCCTTCGGCGAAGCCACCAACATGTCAGGGACAT 975
Db      363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla 377
QY      976 GTTAGGGGCCCATACCTACGCCCTTTGCTGTCGTCGTCGTATGATATTCGGCATCCATAT 1035
Db      378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValargGlyLysCys 396
QY      1036 GATGACCCG-----ACTCGCCCAAGTTATTACAACAATAATTTGCGCAAG 1080
Db      397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
QY      1081 GACTCTGTCATGAGCGCTATCGCGCTCAACATCAAC---TACACCGAGTCCCAATATGAC 1137
Db      417 ProGluValIleGluAlaValGlyAlaGluAlaGlyTyrAspSerCysAsnPheAsp 436

```



```

Db 363 SerAla-----TrrValCysValProAlaSerIleTyrCysAsnAla 377
QY 976 GTGAGGGGCCACTACTACGCCCTTTCGTGCTGCTGTGTATGATATTCGGCAICCATAT 1035
Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396
QY 1036 GATGACCG-----ACTCGCGCAAGTTATTACAACTTTCTGCGCAAG 1080
Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
QY 1081 GACTCTGTCTCAGGAGCTATCGCGCTCAACATCAAC---TACACCCAGTCCAAATAAGAC 1137
Db 417 ProGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436
QY 1138 GTCTACTACGCTTTCAGCAAAACAGCGCACTTTGCTGCGCC-----AATTCATCGAA 1191
Db 437 IleAsnArgAsnPheLeuPheHisGlyAspIrrpMetLysProTyrHisArgLeuValPro 456
QY 1192 GACCTCGAGGAGATCCTGTCTCCCGCTGCTGCTCTCCCTCATCTAT---GGCGACGCC 1248
Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
QY 1249 GATTACATCTCAACTGCTGCTCGCGGCTCAGCGCTTCCCTCGCTGCGCACTATCCCAA 1308
Db 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
QY 1309 GCCGCCAGTTCGGAAGCGCAGGTACACGCCCTGAAAGTC-----AAC 1353
Db 492 GlnAlaGluTyrAlaSerAlaGluLeuGluAspLeuValIleValAspAsnGluHisThr 511
QY 1354 GCGTCGAGTAGNGGGAACATPCGCGAGTATGATGATTTCTCCTCATCTCGCTCTATCAG 1413
Db 512 GlyLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
QY 1414 GCAGGCAATGAGTCCCATACTACACGCCCATCCCTCCCTCGCAATGTTTAAACCGGACT 1473
Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 551
QY 1474 ATCTTCGCT---TGG 1485
Db 552 LeuGlyGlyGluTrp 556

```

RESULT 8

```

US-08-309-341-4
; Sequence 4, Application US/08309341
; Patent No. 5594119
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5594119 No. 5594119 disk of No. 5594119th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,341
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123

```

```

; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Aspergillus Niger
; ORGANISM SOURCE:
; US-08-309-341-4

```

Alignment Scores:

```

Pred. No.: 2,9e-51 Length: 557
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.45% Indels: 76
DB: 1 Gaps: 18
US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-309-341-4 (1-557)

```

```

QY 79 GCAACAATGTACCATCGGTACAGGAACCCGGGCGAGCGGCTGTCCGAGACTACC 138
Db 132 AlalTyrAspLeuArgValLysLysThrAspProSerSerLeuGlyIle-----Asp 148
QY 139 CCGGTGTCAAAATCCTACTCTGGATATGTGCACACTCTCCG---GAGTCCCATACCTTC 195
Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 168
QY 196 TTCGTGCTTCGAGCGCAGACATAACCCAGAACTGCACCTATCACAATGTGTGAT 255
Db 169 TyrTrpPhePheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
QY 256 GGTGGCCCTGGAAGCGATTCTTGTATCGTCTCTTCGGAAGAGTTGGGCCCTTGCATGTC 315
Db 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY 316 AATTCGACTTTTGTATGACTACATCAACCTCATCTCGTGAACAGAGTCTCCAAATTACTA 375
Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
QY 376 TTCCTGTGCCAGCATTCGGGAGTCGGCTTTTCATATAGTATAGCGGTGTGATGGTCCATT 435
Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla-----244
QY 436 AACCTGTAACTGGGGTCTGCGAAAATTTCGAGCTTCGAGAGTTCAGGCGCGGTACCCA 495
Db 244 -----244
QY 496 ACCATTGATGCCACICTGATCATGATACCAATCTGCGCAGAGCGCGTGGAGATC 555
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
QY 556 CTGCAAGCAATCTTAGTGGACTACCTAGCTTGGACTCTAGGTCAGTCTAAGGACTTC 615
Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGluAspPhe 274
QY 616 AGTCTATGGACGAGAGCTATGGAGGCACTATGGTCCIGCATTCCTCAATCATTTTTTAC 675
Db 275 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294
QY 676 GAGCAGAAATCGAGAATTCGCAACCGGTAGTGTAAATGGTTCAGCTTAAATTTCACTCT 735
Db 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
QY 736 CTGGGAATATTACGGCATCATCGACGAGCGGATCCAGGCCCTTACTACCTCGAATC 795
Db 306 ValLeuIleGlyAsnGlyLeuThrGlnTyrGluTyrTyrArgProMet 325
QY 796 GCTGTGAACAATACCTACGCTATCAAGGCTGTCAACGACCGCTCTACACATACATGAAG 855
Db 326 AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla 343

```

```
QY 856 TTTCGCAACCAATGCAATGTTCCAGGATTTCATTTCCACCTGCAACACAAAC 915
Db 344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu 362
QY 916 CGCACCGCATAGCTGACTACGCGCTCTGCGCGAAGCCACCAACATGTGACGGACAAT 975
Db 363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla 377
QY 976 GTTGGAGGGCCATACATACGCTTTGCTGCTGGTGTGTATGATATTCGGCATCATAT 1035
Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396
QY 1036 GATGACCCG-----ACTCCGCCAAGTTATTACACAAATTTCTGGCAAG 1080
Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
QY 1081 GACTCTGTCTGACGACGTATCGCGCTCAACATCAAC---TACACCCAGTCCATAATGAC 1137
Db 417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436
QY 1138 GTCTACTACGCTTCCAGCAAAACAGCGGACTTTGCTGCGCC-----AACTTCATCGAA 1191
Db 437 IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro 456
QY 1192 GACTCTGAGGAGATCCTGCTCTCCCGCTGCTCCCTCACTAT---GGCGACGCC 1248
Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
QY 1249 GATTACATCTGCAACTGGTTCGGCGGTACAGCGGCTTTCCTCGCTGCGGAACTATCCCAA 1308
Db 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
QY 1309 GCCGCCAGTTCGCAACCGCAGGTACACGCCCTGGAAGTC-----AAC 1353
Db 492 GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValValGluAsnGluHisLys 511
QY 1354 GCGCTGAGATGGGAAACTCGCGAGTATGGTAAATTTCTCCTCACTCGCGTCTATGAG 1413
Db 512 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
QY 1414 GCAGGCGCATGAATCCCATCTACACGCCATCCCTCCCTCGCAATGTGTTAACGGGACT 1473
Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 551
QY 1474 ATCTTCGGT---TGG 1485
Db 552 LeuGlyGlyGluTrp 556

RESULT 9
US-08-608-267-4
; Sequence 4, Application US/08608267
; Patent No. 5688663
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5688663o No. 5688663disk of No. 5688663th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,267
; FILING DATE: 28-FEB-1996
```

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowrey, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 557 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Aspergillus Niger
US-08-608-267-4

Alignment Scores:
Pred. No.: 2,9e-51
Score: 623.00
Percent Similarity: 48.25%
Best Local Similarity: 31.96%
Query Match: 21.45%
Gaps: 1
DB: 18

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-267-4 (1-557)
QY 79 GCAACAAATGTCACATCCCGTACAGGAACCCGGGCGAGAGGGCGTCTGCGAGACTACC 138
Db 132 AlaTyrAspLeuArgValLysThrAspProSerSerLeuGlyIle-----Asp 148
QY 139 CCGGGTGTCAATCTCTACTCTGGATATGTCACACCTCTCCC---GAGTCCCATACCTTC 195
Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAsnGluAsnAspHisLeuPhe 168
QY 196 TTCGTGTCTTCGAAGCCAGACATAACCCAGAACTGCACCTATCACATTTGTGGTGAAT 255
Db 169 TyrTrpPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
QY 256 GGTGCGCCCTGGAAGCGATTCTTTGATCGGTCTCTCGAAGAGTTCGGCGCTTGCATGTC 315
Db 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSerIle 208
QY 316 AATTCGACITTTGATGACTACATCAACCTCTACTCTGGAACGGTCTCCCAATTTACTA 375
Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
QY 376 TTCCTGTCCCGCCATTCGGGATCGGCTTTTCATATAGTATGATACGCTTGTGATGGTCCATT 435
Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla-----244
QY 436 AACCTGTAACTGGGCTGCTCGAAATTCGAGCTTTGAGAGGTTCAGGCGCGGTACCCA 495
Db 244 -----244
QY 496 ACCATTGATGCCACTCTGATCATCTACCATCTTCGCGAGAGGCCCGCTTGGGAGATC 555
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
QY 556 CTGCAAGGATTCCTTAGTGGACTACTAGCTTTGGACTCTTAGGGTGCAGCTCAAGGACTTC 615
Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274
QY 616 AGTCTATGACGCGAGAGCTATGAGGGCACTATGTCCTGCTGCAATTCATCATTTTAC 675
Db 275 HisIleAlaGlySerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294
QY 676 GAGCAGAAATGAGAGATTCGCCAAGGTAGTGTATGTTGTTTCACTTCACTCT 735
```

295 SerHisLysArg-:::|||||
736 CTGGGAATATTAAACGATCATCGACGAGGCGATCCAGGCCCTTACTACCTCGAATC 795
306 ValLeuLeuGlyAsnGlyLeuThrAspGlyLeuThrGlnTyrTyrArgProMet 325
796 GCTGTGAACAACTACGCTATCAAGGCTGTCAACGAGCGCTCTCAACTACATGAAG 855
326 AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla 343
856 TTTCGCAACAAATGCCAATGTTGCCAGGATTGATTCACCTCGCAACACAGCAAC 915
344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerGlu 362
916 CGCACCGCATTAGCTGACTACGCTCGCTCGCGCAACCCACACATGTCGAGGACAAT 975
363 SerAla-----TrpValCysValProAlaSerValIleTyrCysAsnAla 377
976 GTTAGGGGCCATPACTACGCTTTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1035
378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396
1036 GATGACCGC-----ACTCGCCCAAGTTATTACAAATTTCTGGCAAG 1080
397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
1081 GACTCTGTGATGACGCTATCGGCGTCACATCAAC---TACACCCAGTCCATATGAC 1137
417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnHeAsp 436
1138 GTCTACTAGCTTTCCAGCAACAGCGACTTGTCTGTCGCCC-----AACTTCATCGAA 1191
437 IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro 456
1192 GACCTCGAGAGATCTTCTCTCCCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1248
457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
1249 GATTACATCTGCACTGCTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCGGTGTCGCGGT 1308
472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
1309 GCGGCGCATGTCGGAAGCGACGAGGTCACGCGCCCTGAAAGTC-----AAC 1353
492 GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValValGluAsnGluHisLys 511
1354 GCGCTCGAGTATGGGAACCTCGGAGATGATGTAATTTCTCTCTCTCTCTCTCTCTCTCT 1413
512 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
1414 GCAGGCCCATGAAGTCCCATCTACAGCCCATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1473
532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPheAsnArgTrp 551
1474 ATCTTCGGT---TGG 1485
552 LeuGlyGlyGluTrp 556

RESULT 10
US-08-608-452-4
; Sequence No. 5693510
; Patent No. 5693510
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 56935100 No. 56935100disk of No. 5693510th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York

COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/608,452
FILING DATE: 28-FEB-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,341
FILING DATE: 20-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowney, Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4247,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 557 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORGANISM: Aspergillus Niger
US-08-608-452-4
Alignment Scores:
Pred. No.: 2,9e-51 Length: 557
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.45% Indels: 76
DB: 1 Gaps: 18
US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-608-452-4 (1-557)
QY 79 GCACAAATGTCACCATCGGTACAGAACCGGGGACAGAGCGGTCTGCGAGACTACC 138
Db 132 AlaTyrAspLeuArgValLysThrAspProSerSerLeuGlyIle-----Asp 148
QY 139 CGGGGTGCAATCCCTACTCTGATATGTCGACACCTCTCCC---GAGTCCCATACCTTC 195
Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAspAsnGluAsnAspLysHisLeuPhe 168
QY 196 TTCTGTTCTTCCAGCCAGACATAACCCAGAACTGCACCTATCATTTGGTTGAAT 255
Db 169 TyrTrpPhePheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
QY 256 GGTGCGCCCTGGAAGCCATCTTTGATCGTCTCTTCGAGAGAGTGGGCCCTTGCCTGTC 315
Db 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGlyLeuGlyProSerSerIle 208
QY 316 AATCGACTTTTGTGATGACTACATCACTCGTGAACGAGGTCTCTCAATTACTA 375
Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
QY 376 TTCTGTCTCCCGCATTTGGAGTCTGCTTTTCATATAGTATGATGATGATGGTGGTCCATT 435
Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla-----244
QY 436 ACCCTGTAACTGGGTGCTCGAATAATTCGAGAGTTCGAGGAGTTCAGGCGCGGTACCCA 495
Db 244 -----244
QY 496 ACCATTGATGCCACTCTGATGATGACTACCACTACCACTGTCGCGGACAGCGCGTGGGAGATC 555
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258

Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla----- 244
QY 436 AACCCITGTAACCTGGGGTCGTCGAAATTCGAGCCTTCAGAGTTCAGGCCGGTACCCA 495
Db 244 ----- 244
QY 496 ACCATTGATCCACTCTGATCATACCTACCAATCTGCCGAGAGCGCGTGGGAGATC 555
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValTyrAlaLeu 258
QY 556 CTCGAGGATTCCTAGTGGACATACCTAGTCTGGAGCTCTAGGGTCAGCTCAAGACTTC 615
Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274
QY 616 AGTCTATGGACGAGGAGTATGGAGCGCACTATGGCTCATCTCTTCATCATCTTTAC 675
Db 275 HisIleAlaGlySerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294
QY 676 GAGCAGAATGAGAGAATGCCAACGGTAGTGTAAATGGTTCAGCTTAATTTCAACTCT 735
Db 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
QY 736 CTGGGAATTAATAACGGCATCATCGAGGCGGATCCAGGCCCTTACTACCTGAATTC 795
Db 306 ValLeuIleGlyAsnGlyLeuThrAspGlyLeuThrGlnTyrGluTyrTyrArgProMet 325
QY 796 GCTGTGCAACAATACCTACGCTATCAAGCGTGTCAAGCAGAGCCCTACACTACATGAAG 855
Db 326 AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla 343
QY 856 TTTCGCCAACAATGCCAATGTTGCCAGGATTTGATTTCCACCTGCCAACAGACAAAC 915
Db 344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu 362
QY 916 CGCAGCGCATAGCTACCTACGCCCTCTCGCGGAGAGCCACCACCAATGTCAGGGCAAT 975
Db 363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAsnAla 377
QY 976 GTTAGGGGCCATACCTACGCCCTTGTGCTGTGGTGTATCATATTCGGATCCATAT 1035
Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgLysCys 396
QY 1036 GATGACCCG-----ACTCGGCCAAGTATTACACAAATTTCTGGCAAG 1080
Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
QY 1081 GACTCTGTGATGAGCTATCGCGCTCAACATCAC---TACACCCAGTCCCAATAATGAC 1137
Db 417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436
QY 1138 GTCTACTACGCTTTCAGCAACAGCGGACTTGTGTGCGCC-----AACTTCATCGAA 1191
Db 437 IleAsnArgAsnPheLeuPheHisGlyAspTyrMetLysProTyrHisArgLeuValPro 456
QY 1192 GACCTCGAGGAGATCCTGCTCTCCCGGCGGTGTCTCCCTCATCTAT---GGCAGCGCC 1248
Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
QY 1249 GATTACATCTGCACTGCTCGCGCGTCAAGCGCGTTCCTCGCTGCGCAACTCCCAA 1308
Db 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpGly 491
QY 1309 GCGCCCGAGTTCGGAAGCGGGGACAGCGCCCGCGGAAAGTC-----AAC 1353
Db 492 GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValGluAsnGluHisLys 511
QY 1354 GCGTCGAGATATGGGAACCTCGGAGTATGTTATTTCTCTTCTACCTCGGCTCATGAG 1413
Db 512 GlyLysLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
QY 1414 GCAGGCCATGAGTCCCATATACAGCCCATCGGCTCCCTGCAATTTGTTAAACCGGACT 1473
Db ----- 1473

Db 532 GlyGlyHisMetValProMetAspGlnProGluSerSerLeuGluPhePheAsnArgTrp 551
QY 1474 ATCTTCGGT---IGG 1485
Db 552 LeuGlyGlyGluTrp 556
RESULT 12
US-08-967-149-4
; Sequence 4, Application US/08967149
; Patent No. 5939305
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie Sue
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENE ENCODING CARBOXYPEPTIDASE OF
; TITLE OF INVENTION: ASPERGILLIUS NIGER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5939305 No. 5939305disk of No. 5939305th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,452
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 08/309,341
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4247.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus Niger
US-08-967-149-4
Alignment Scores:
Pred. No.: 2,9e-51 Length: 557
Score: 623.00 Matches: 155
Percent Similarity: 48.25% Conservative: 79
Best Local Similarity: 31.96% Mismatches: 175
Query Match: 21.45% Indels: 76
DB: 2 Gaps: 18
US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-967-149-4 (1-557)
QY 79 GCAACAATGTCAACCATCGGTACAGNACCCGGGAGAGGGCTGCGAGACTACC 138
Db 132 AlaTyrAspLeuArgValLysLysThrAspProSerSerLeuGlyIle-----Asp 148
QY 139 CCGGTGTCAATCTACTCTGATGTAIGTCGACACTCTCCG---GAGTCCCATACCTTC 195
Db 149 ProGlyValLysGlnTyrThrGlyTyrLeuAspAsnGluAsnAspLysHisLeuPhe 168

```
QY 196 TTCGGTCTTCGAGCGACAGATACCAAGAACTGCACCTATACATTTGGTTGAAT 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 TyrTrpPheGluSerArgAsnAspProGluAsnAspProValValLeuTrpLeuAsn 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 256 GGTGCGCCCTGGAAGCGATTCCTTTGATCGGTCTCTTCGAGAGATGGCCCTTGCACATGC 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 GlyGlyProGlyCysSerSerLeuThrGlyLeuPheMetGluLeuGlyProSerSertile 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 AATTCGACTTTTTCATGACTACATCAACCTCCTACCTCGTGGAGAGAGCTGCCAATTTACTA 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 AsnLysLysIleGlnProValTyrAsnAspTyrAlaTrpAsnSerAsnAlaSerValIle 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 376 TTCCTGTCCAGCCATTCGGAGCTCGCTTTTCATATAGTACATACGGTTGATGGTCCATT 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 PheLeuAspGlnProValAsnValGlyTyrSerTyrSerAsnSerAla 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 436 AACCCGTGAATGGGGTGGTGCAGAAATTCGAGAGTTCAGGCGCGGTACCCA 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 244 ----- 244
QY 496 ACCATGATGCCACTCTGATCATACTACCAATCTGCGCGAGAGCGCTTGGGAGATC 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 -----ValSerAspThrValAlaAlaGlyLysAspValIleAlaLeu 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 556 CTCGAAGGATTCCTTAGTGGACTTACCTAGCTTGGACTCTAGGCTGAGCTTAAGGACTTC 615
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 LeuThrLeuPhePheLysGlnPhePro-----GluTyrAlaLysGlnAspPhe 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 616 AGTCTAAGCAGCGAGGACTATGAGGCGCACTATGCTCCTGCATCTTCAATCATTTTAC 675
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 HisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAlaSerGluIleLeu 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 676 GAGCAAGATGAGAGATTCGCAACGGTAGTGTAAAGGTCTTCAGCTTAATTTCAACTCT 735
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 SerHisLysLysArg-----AsnIleAsnLeuGlnSer 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 736 CTGGGAATTAATTAACGGCATATCGAGCGGATCCAGCGCCCTTACTACCTGATTC 795
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 ValLeuIleGlyAsnGlyLeuThrAspGlyLeuThrGlnTyrGlyTyrArgProMet 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 796 GCTGTGAACAAWACCTACGGTATCAGGCTGTCACGAGCGCTACACTACATCAAG 855
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 AlaCysGlyAspGly---GlyTyrProAlaValLeuAspGlu---GlySerCysGlnAla 343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 856 TTTCGCAACCAATATGCAATGTTGCCAGATTTGATTCACCTGCAACAGACAAAC 915
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 344 MetAspAsnAlaLeuProArg---CysGlnSerMetIleGluSerCysTyrSerSerGlu 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 916 CGCACCGCATTAGCTGACTACGCTCTGCGCCGCAAGCCACCAATGTCAGGAGCAAT 975
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 SerAla-----TrpValCysValProAlaSerIleTyrCysAsnAla 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 976 GTTCAGGGGCCATACACTACGCTTTCGCTGCTGGTGTGTATGATATTCGGCATCCATAT 1035
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 378 LeuLeuAlaProTyrGlnArg---ThrGlyGlnAsnValTyrAspValArgGlyLysCys 396
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1036 GATGACCGC-----ACTCCGCCAAGTATTACACAAATTTCTGGCAAG 1080
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 GluAspSerSerAsnLeuCysTyrSerAlaMetGlyTyrValSerAspTyrLeuAsnLys 416
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1081 GACTCTGTATGACGCTATCGCGCTCAACATCAAC---TACACCCAGTCCCAATATGAC 1137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 ThrGluValIleGluAlaValGlyAlaGluValAsnGlyTyrAspSerCysAsnPheAsp 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1138 GTCTACTACCTTTCACAGCAACAGCGACTTTCCTGCGCC-----AACTTCATCGAA 1191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 437 IleAsnArgAsnPheLeuPheHisGlyAspTrpMetLysProTyrHisArgLeuValPro 456
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1192 GACTTCGAGGAGATCTTGTCTCCCGGTGGTGTCTCCCTCATCTAT---GGCAGCGCC 1248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 GlyLeuLeuGlu-----GlnIleProVal-----LeuIleTyrAlaGlyAspAla 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1249 GATTACATCTGCACAACTGGTTCGGCGGTTCAGGCGGTTCCCTCGCTCGCACTACTCCAA 1308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 472 AspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrGluAlaLeuGluTrpProGly 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1309 GCGCGCCACTTCGAGCGCAGGTACACGCCCTCGAAAGTC-----AAC 1353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 492 GlnAlaGluTyrAlaSerAlaLysLeuGluAspLeuValValGluAsnGluHisLys 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1354 GCGCTCGAGTATGGGAAACTCGCGAGTATGTAATTTCTCTTCACTCGCGTCTATGAG 1413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 GlyLysIleGlyGlnValLysSerHisGlyAsnPheThrPheMetArgLeuTyrGly 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1414 GCAGGCCATGAAGTCCCATCTACACGACCATCGCTCCCTGCAATTTGTTAACCGGACT 1473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 532 GlyGlyHisMetValProMetAspGlnProGluSerLeuGluPheAsnArgTrp 551
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1474 ATCTTCGGT---TGG 1485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 LeuGlyGlyLeuTrp 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-640-305-4
; Sequence 4, Application US/09640305
; Patent No. RE37447
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/640,305
; FILING DATE: 16-Aug-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE: 06-FEB-1995
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 491 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-640-305-4

Alignment Scores:
Pred. No.: 4.81e-46 Length: 491
Score: 568.50 Matches: 151
Percent Similarity: 44.65% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 184
```


Query Match:	19..58%	Indels:	85
DB:	1	Gaps:	14
US-09-712-338-1_COPY_55_1662 (1-1608) x US-09-640-305-4 (1-491)			
QY	79	GCAACAACTGTCACCATCGGTACAGGAACCCGGGCGAGGGCGTCTGCGAGACTACC	138
Db	62	AlaIyrSerLeuArgIleIysProIeuAspProIysSerLeuGlyVal	77
QY	139	CGGGTGTCAAAATCCTACTCTCGATATGTCGACACCTCTCCGAGTCCCATACCTTCCTC	198
Db	78	AspThrValIysGlnIrpSerGlyIyrIeuAspIyrGlnAspSerIyshiPhePheIyr	97
QY	199	TGGTTCTTCGAGCCAGACATTAACCCGAAACTGCACCTATCACCATTCTGGTGTGAATGGT	258
Db	98	TrpPhePheGluSerArgAsnAspProGluAsnAspProValIleIeuTrpIeuAsnGly	117
QY	259	GGCCCTGGAAGCGATCTTTTGATCGTCTCTTCGAGAGTGGGGCCCTGCATGTCATCAAT	318
Db	118	GlyProGlyCysSerSerPheValGlyPhePheGluLeuGlyProSerSerIleGly	137
QY	319	TGCACTTTTGATGACTACATCAACCTCACTCGTGGAAACGAGGTCTCCAAATTACTATTC	378
Db	138	AlaAspLeuIysProIleIyrAsnProIyrSerIrpAsnSerAsnAlaSerValIlePhe	157
QY	379	CTGTCCCGACCATGGAGTCGGCTTTTCATATAGTAGTATCGGTTGATGGTCCATTAAAC	438
Db	158	LeuaspGlnProValIgyValIgyPheSerIyrGlyAsp	170
QY	439	CCTGTAACTGGGGTGGTGCAGAAATTCGAGCTTTCGAGGAGTTCAGGGCCGGTACCAACC	498
Db	170	-----	170
QY	499	ATTGATGCCACTCTGATCGATACCTACCAATCTGCCGACAGGGCGCTTGGGAGATCCTG	558
Db	171	-----SerIyshiValSerThrThrAspAlaAlaIysAspValIyrIlePheLeu	187
QY	559	CAAGGATTCCTTAGTGGACTACCTAGCTTGACTCTAGGGTGCAGTCTAAGGACTTCAGT	618
Db	188	AspLeuPhePheGluArgPheProHisLeu	203
QY	619	CTATGGACGGAGACTATGGAGGSCACTATGGTCTCTGCAATCTCTCAATCAITTTTACGAG	678
Db	204	IleSerGlyGluSerIyrAlaGlyHisIyrLeuProIysIleAlaHis	219
QY	679	CAGAAATCAGAGAATTGCCACGGTAGCTGTAAAGTGTCTTCAGCWTATTTCAACTCTCTG	738
Db	220	-----GluIleAlaValHisAlaGluAspSerSerPheAsnLeuSerSerVal	236
QY	739	GGAATATTAAACGGCATCATCGACGAGGCGATCCAGGCCCTTACTACCTGAATTCGCT	798
Db	237	LeuIleGlyAsnGlyPheThrAspProLeuThrGlnIyrGlnIyrGluProMetAla	256
QY	799	GTCAACAATACCTACGGTATCAGGCTGTCAACGAGACCGCTCTACAACTACATGAAGTTT	858
Db	257	CysGlyGluGly---GlyIyrProAlaValLeuGlu	267
QY	859	GCCAAACCAATGCCAATGGTTCGCCAGGAT	888
Db	268	-----ProGluAspCysLeuIleuMetAsnArgAsnLeuProLeuCysLeuSer	283
QY	889	TGTATTTCCACCTGCAACACAGACAAACCGCACCGCATTAGCTACAGCCCTCTCGGCC	948
Db	284	LeuValAspArgCysIyrIysSerHis	298
QY	949	GAAGCCACCAACATGTCAGGACAAATGTTGAGGGGCCATACCTACGCCCTTGTGTCGT	1008
Db	299	LeuAlaAspArgIyrCysGluGlnIleThrGly---ValIyrGluIysSerGlyArg	317
QY	1009	GGTGTGTATGATATTCGG	1047
Db	318	AsnProIyrAspIleArgSerIysCysGluAlaGluAspSerGlyAlaCysIyrGln	337

QY	1048	CGCCCAAGTATTACAAACAATTTCTGGCAAGAGACTCTGTCTATGGACGCTATCGCGTGC	1107
Db	338	GlUGluIleTyrIleSerAspTyrLeuAsnGlnGluValGlnArgAlaLeuGlyThr	357
QY	1108	RACATCAACTACACCCACTCC---AATAATGACGCTCTACTACGCTTTCCAGCAACAGGC	1164
Db	358	AspValSerSerPheGlnGlyCysSerSerAspValGlyIleGlyPheAlaPheThrGly	377
QY	1165	GACITGTGCTGGCCCAACTTCATCAAGACACTCGAGGAGATCCTTGTCTCTCCCGTGGT	1224
Db	378	AspGlyProSerPro---PheHisGluTyrValAlaGluLeuLeuAspGlnAspIleAsn	396
QY	1225	GTCTCCCTCATCTATGGCGAGCGCGCATTACATCTGCAACTGTGTGGCGGTCAAGCCGT	1284
Db	397	ValLeuIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGlyAsnLeuAlaTrp	416
QY	1285	TCCCTGGTGGAACTACTCCCAAGCGCCAGTTCGGAGCGCAGGATACAGCCCGCTG	1344
Db	417	ThrGluLysLeuGluTrpArgTyrAsnGluGluTyrLysGlnValLeuArgThrTrp	436
QY	1345	AAAGTCAACGCGCTCGAG-----TATGGGAACACTCGGAGTAGTGGTAATTCCTC	1398
Db	437	LysSerGluGluThrAspGluThrIleGlyGluThrLysSerTyrGlyProLeuThrTyr	456
QY	1399	ACTCGCTCATGAGCAGCGCCATGAAGTCCCATACTACCAAGCCCATCGCTCCCTGCAA	1458
Db	457	LeuArgIleTyrAspAlaGlyHisMetValProHisAspGlnProGluAsnSerLeuGln	476
QY	1459	TGTGTTAACGGACTATC	1476
Db	477	MetValAsnSerTrpIle	482
RESULT 14			
US-08-360-673-4			
; Sequence 4, Application US/08360673			
; Patent No. 5679544			
; GENERAL INFORMATION:			
; APPLICANT: Fleer, Reinhard			
; APPLICANT: Fournier, Alain			
; APPLICANT: Veh, Patrice			
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR			
; TITLE OF INVENTION: PREPARATION AND USE			
; NUMBER OF SEQUENCES: 17			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Rhone-Poulenc Roter Inc.			
; STREET: 500 Arcola Rd. 3043			
; CITY: Collegeville			
; STATE: PA			
; COUNTRY: USA			
; ZIP: 19002			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/360,673			
; FILING DATE:			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: WO PCT/FR93/00623			
; FILING DATE: 23-JUN-1993			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: FR 92/07785			
; FILING DATE: 25-JUN-1992			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Smith, Julie K.			
; REGISTRATION NUMBER: 38,619			
; REFERENCE/DOCKET NUMBER: ST92040-US			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (610)454-9839			
; TELEFAX: (610)454-3808			
; INFORMATION FOR SEQ ID NO: 4:			

SEQUENCE CHARACTERISTICS:
LENGTH: 491 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-360-673-4

Alignment Scores:
Pred. No.: 491
Score: 568.50
Percent Similarity: 44.65%
Best Local Similarity: 31.07%
Query Match: 19.58%
DB: 14

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-360-673-4 (1-491)

QY 79 GCAACAAATGTCACCATCGGTACAGGAACCGGGGAGAGGGCTCTCGAGACTACC 138
Db 62 AlaTyrSerLeuArgIleLysProLeuAspProLysSerLeuGlyVal----- 77
QY 139 CGGGTGTCAATCTACTGCTGATGATGCGACACCTCTCCCGAGTCCCATACCTCTTC 198
Db 78 AspThrValLysGlnTrpSerGlyTyrLeuAspTyrGlnAspSerLysHisPheTyr 97
QY 199 TGGTCTTCGAGCCAGCATACACCCAGAACTGACCTATCATCATTTGGTGAATGGT 258
Db 98 TrpPheGluSerArgAsnAspProGluAsnAspProValIleLeuTrpLeuAsnGly 117
QY 259 GGCCTCGGAGGATCTTTGATCGCTCTCTCGAAGAGTGGGCGCTTCGCCATGTCAT 318
Db 118 GlyProGlyCysSerPheValGlyLeuPheGluLeuGlyProSerSerIleGly 137
QY 319 TCAGCTTTGATGACTACATCAACCTCTACTCTGGAACGAGTCTCCATTTACTATTC 378
Db 138 AlaAspLeuLysProIleTyrAsnProTyrSerTrpAsnSerAsnAlaSerValIlePhe 157
QY 379 CTCTCCAGCATTGGAGTGGCTTTTCATATAGTACATCGGTTGATGGTCCATTAAC 438
Db 158 LeuAspGlnProValGlyValGlyPheSerTyrGlyAsp----- 170
QY 439 CTTGTAACGGGTGCTCGAAATTCGAGTTCGAGGAGTTCAGGGCGGTACCCCAACC 498
Db 170 ----- 170
QY 499 ATTGATCCCACTGTGATCGATACCTACCAATCTCCCGCAGAGCGGCTTGGGAGATCTG 558
Db 171 -----SerLysValSerThrThrAspAspAlaAlaLysAspValTyrIlePheLeu 187
QY 559 CAAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGGTGCAGTCTAAGGACITCAGT 618
Db 188 AspLeuPheGluArgPheProHisLeu-----ArgAsnAspPheHis 203
QY 619 CTATGGAGGAGAGCTATGAGGCGCATATGGTCTCGATCTTCAATCAATTTACGAG 678
Db 204 IleSerGlyGluSerTyrAlaGlyHisTyrLeuProLysIleAlaHis----- 219
QY 679 CAGATGAGAGAAATGCCAGGTAGTCTTAATGTTGCTCAGCTTAATTCACATCTCTG 738
Db 220 -----GluIleAlaValValHisAlaGluAspSerSerPheAsnLeuSerVal 236
QY 739 GGAATTTATACGGCATCATCGAGGCGCATCCAGCCCTTACTACCTCAATTCGCT 798
Db 237 LeuIleGlyAsnGlyPheThrAspProLeuThrGlnTyrGlnTyrGluProMetaIa 256
QY 799 GTGACAAATACCTAGGGTATCAGGCTGTCAACGAGACCGTCTACACTACATGAGTTT 858
Db 257 CysGlyGluGly---GlyTyrProAlaValLeuGlu----- 267
QY 859 GCCAACCAATGCCAAATGGTTCACGAGT----- 888
Db 268 -----ProGluAspCysLeuAspMetAsnArgAsnLeuProLeuCysLeuSer 283

QY 889 TTGATTTCCACTGCAACAGACAAACCGCACCGCATTTAGCTACGCCCTCTGCGCC 948
Db 284 LeuValAspArgCysTyrLysSerHis-----SerValPheSerCysVal 298
QY 949 GAAGCCACCAACATGTCAGGACACATGTTAGGGGCCCATACACCTTTGCTGTGCTG 1008
Db 299 LeuAlaAspArgTyrCysGluGlnIleThrGly---ValTyrGluLysSerGlyArg 317
QY 1009 GGTGTGTATGATATCGG-----CATCCATATGATGACCCG-----ACT 1047
Db 318 AsnProTyrAspIleArgSerLysCysGluAlaGluAspAspSerGlyAlaCysTyrGln 337
QY 1048 CGCCCAAGTTATACACAAATTTCTGCGCAAGAGCTCTGTATGACGCTATCGCGGTC 1107
Db 338 GluGluIleTyrIleSerAspTyrLeuAsnGlnGluValGlnArgAlaLeuGlyThr 357
QY 1108 AACATCAACTACACCCAGTCC---AATAATGAGTCTACTACGCTTTCCAGCAACAGCC 1164
Db 358 ASPValSerSerPheGlnGlyCysSerSerAspValGlyIleGlyPheAlaPheThrGly 377
QY 1165 GACTTTGTCGCGCCCACTTCATCGAAGACCTCGAGAGATCCTTGTCTCTCCCGTGGT 1224
Db 378 AspGlyProSerPro---PheHisGlnTyrValAlaGluLeuLeuAspGlnAspIleAsn 396
QY 1225 GTCTCCTCATCTATGCGGACGCGGATTACATCTGCAACTGGTTCGGCGGTCAAGCGGT 1284
Db 397 ValIleIleTyrAlaGlyAspLysAspTyrIleCysAsnTrpLeuGlyAsnLeuAlaTrp 416
QY 1285 TCCTCTGCTCGAACTACTCCCAAGCGCGGCGGTCGGAAGCGGCGGTACACGCCCTG 1344
Db 417 ThrGluLysLeuGlnTrpArgTyrAsnGluLysLysGlnValLeuArgThrTrp 436
QY 1345 AAGTCAACCGGCTCGAG-----TATGGGAACTCGCGAGTATGTAATTTCTCCTC 1398
Db 437 LysSerGluGluThrAspGluThrIleGlyLeuThrLysSerTyrGlyProLeuThrTrp 456
QY 1399 ACTCGGCTCTATGAGCGAGCGCATGAAGTCCCATCTACCAAGCCCATCGCTCCCTGCA 1458
Db 457 LeuArgIleTyrAspAlaGlyHisMetValProHisAspGlnProGluAsnSerLeuGln 476
QY 1459 TTGTTTAAACGGACTATC 1476
Db 477 MetValAsnSerTrpIle 482

RESULT 15
US-08-899-324-33
Sequence 33, Application US/08899324
Patent No. 5945329
GENERAL INFORMATION:
APPLICANT: Bredam, Klaus
APPLICANT: Keilland-Brandt, Morten
APPLICANT: Mortensen, Uffe
APPLICANT: Olesen, Kjeld
APPLICANT: Stennicke, Henning
APPLICANT: Wagner, Fred
TITLE OF INVENTION: CUSTOMIZED PROTEASES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 3100 No. 5945329west Center, 90 S. 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: U.S.A.
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/08/899,324
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/329,892
 ; FILING DATE: 27-OCT-1994
 ; APPLICATION NUMBER: 08/144,704
 ; FILING DATE: 28-OCT-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kettleberger, Denise M
 ; REGISTRATION NUMBER: 33,924
 ; REFERENCE/DOCKET NUMBER: 8648.44USC1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 612/332-5300
 ; TELEFAX: 612/332-9081
 ; TELEX:

; INFORMATION FOR SEQ ID NO: 33:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 532 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: Internal
 ; ORIGINAL SOURCE:
 ; US-08-899-324-33

Alignment Scores:
 Pred. No.: 1,1e-41 Length: 532
 Score: 523.50 Matches: 147
 Percent Similarity: 43.68% Conservative: 74
 Best Local Similarity: 29.05% Mismatches: 200
 Query Match: 18.03% Indels: 85
 DB: 2 Gaps: 19

US-09-712-338-1_COPY_55_1662 (1-1608) x US-08-899-324-33 (1-532)

QY 43 CCCAAGAACCCACCGGGGTCAAGACT-----CITACAAACCCCAACAACT 87
 DB 82 ProLysPheProGluAlaIleLysThrLysLysAspTrpAspPheValValLysAsnAsp 101
 QY 88 GTCACCATCCGGTAC-----AAGGAACCCGGGGCAGAGGCGTC 126
 DB 102 AlaIleGluAsnTrpGlnLeuArgValAsnLysIleLysAspProLysIleLeuGlyIle 121
 QY 127 TCGAGACTACCCGGGTGCAAACTCCACTGCTGGATATGTCGACACCTCCCGAG--- 183
 DB 122 -----AspProAsnValThrGlnTrpGlyThrGlyThrLysValGluAspGluAsp 138
 QY 184 TCCCATACCTCTCTCTGCTTCGAGCCAGACATAACCCAGAACTGCACCTATCACA 243
 DB 139 LysHisPhePheThrPheGluSerArgAsnAspProAlaLysAspProValIle 158
 QY 244 TTGGTGTGAATGTCGCCCTGGAAGCGATTCCTTGATCGTCTCTTCGAAAGAGTTGGGC 303
 DB 159 LeuTrpLeuAsnGlyLysProGlyCysSerSerLeuThrGlyLeuPheGluLeuGly 178
 QY 304 CTTTGGCATGTCAATTCGACTTTTTCGACATACATCAACCTCACTCGTGTGAAACGAGTC 363
 DB 179 ProSerSerIleGlyProAspLeuLysProIleGlyAsnProTrpSerTrpPheSerAsn 198
 QY 364 TCCAAATTACTATTCCTGTCGCCACCATGTCGAGTCGGCTTTTCATATAGTAGTACGGTT 423
 DB 199 AlaThrValIlePheLeuAspGlnProValAsnValGlyPheSerTrp----- 215
 QY 424 GATGGGTCCATTACCCCTGTAACCTGGGGTCTCGCAAAATCGAGCTTTGCGAGGAGTTCAG 483
 DB 216 -----GlySerSer 218
 QY 484 GCGCGGTACCCACCACTTATGATGCCACTCTGATCGATACCACTTCGCCGAGAGGCC 543
 DB 219 Gly-----ValSerAsnThrValAlaAlaGlyLysAsp 229
 QY 544 GCTTGGGAGATCTCGAAGGATTCCTTAGTGGACTACCTAGCTTGGACICTAGGGTGCAG 603

DB 230 ValTyrAsnPheLeuGluLeuPheAspGlnPheProGluTyrValAsnLysGlyGln 249
 QY 604 TCTAAGGACTTCAGTCTATGGAGGAGAGCTATGAGGAGCAGCTATGCTCTGCATTCTTC 663
 DB 250 -----AspPheHisIleAlaGlyGluSerTyrAlaGlyHisTyrIleProValPheAla 267
 QY 664 AATCAATTTTACGAGCAGAAATGAGAGAAATGCCCAACGGTAGTGTAAIGSTGTTTCAGCTT 723
 DB 268 SerGluIleLeuSerHisLysAspArg----- 276
 QY 724 AATTCAACTCTCTGGGAATTATT-----AACGGCATCATCGACAGGCGATCCAGGCC 777
 DB 277 AsnPheAsnLeuThrSerValIleLeuGlyAsnGlyLeuThrAspProLeuThrGlnTyr 296
 QY 778 CTTTACTACCTGAATTGCTGTGAACAATACCTAGGTATCAAGCTGTCACAGGAGACC 837
 DB 297 AsnIleTyrGluProMetAlaCysGlyGluGly---GlyGluProSerVal----- 312
 QY 838 GCTTCACTACATGATGAAGTTTGCACCAATGCCAAATGGTTCGCCAGGATTTGATTTC 897
 DB 313 -----LeuProSerGluGluCysSerAlaMetGluAspSerLeuGlu 326
 QY 898 ACCTGCAAAACAGACAAACCGCACCTTACCTAGCTACCTACGCCCTC-----TGCGCCGAA 951
 DB 327 ArgCysLeuGlyLeuIleGluSerSerTyrAspSerGlnSerValTrpSerCysValPro 346
 QY 952 GCCACCAACAATTCCTGCCAAGGACTCTGTATGAGGGCCATACCTAGCCCTTCTGCTGCTGGT 1011
 DB 347 AlaThrIleTyrCysAsnAsnAlaGlnLeuAlaProTyrGlnArg---ThrGlyArgAsn 365
 QY 1012 GTGTATGATATTCGGCATCCATATGATGAC-----CCGACTCCGCCAAGT 1056
 DB 366 ValTyrAspIleArgLysAspCysGluGlyGlyAsnLeuCysTyrProThrLeuGlnAsp 385
 QY 1057 TATTACACAAATTCCTGCCAAGGACTCTGTATGAGGCGCATCGCGCTCAACATCAAC 1116
 DB 386 ---IleAspAspTyrIleAsnGlnAspTyrValLysGluAlaValGlyAlaGluValAsp 404
 QY 1117 TACACCCAGTCC---AAVAAATGACGTCTACTACGCTTCCACCAACAGCGGACITTC 1173
 DB 405 HisTyrGluSerCysAsnPheAspIleAsnArgAsnPheLeuPheAlaGlyAspTrpMet 424
 QY 1174 TGGCCCACTTTCATCGAAGACCTCGAGGAGATCCTTGTCTCTCCCGTGGGTGTCTCCCTC 1233
 DB 425 LysPro---TyrHisThrAlaValThrAspLeuLeuAsnGlnAspLeuProIleLeuVal 443
 QY 1234 ATCTATGGCGAGCCGATACATCTGCAACGTGTCGGGGTCAGCGCGTTCCTCGCT 1293
 DB 444 TyrAlaGlyAspLysAspPheIleCysAsnTrpLeuGlyAsnLysAlaTrpThrAspVal 463
 QY 1294 GCGAACTACTCCCAAGCGCCGAGTTCGGAAGC-----GCAGGTACAGCCCGCTG 1344
 DB 464 LeuProTrpLysTyrAspGluPheAlaSerGlnLysValArgAsnTrpThrAlaSer 483
 QY 1345 AAAGTCACGGCGTCGAGTATGGGAAACTCCGAGTATGGTAATTCCTCTTCATCTGC 1404
 DB 484 IleThrAspGluValAla---GlyGluValLysSerTyrLysHisPheThrTyrLeuArg 502
 QY 1405 GTCTATGAGGCGGATGAGTCCATCTACAGCCCATCAGCGCGTTCCTCGCTGCAATGTGTT 1464
 DB 503 ValPheAsnGlyGlyHisMetValProPheAspValProGluAsnAlaLeuSerMetVal 522
 QY 1465 AACCGGACTATCTTCGGT 1482
 DB 523 AsnGluTrpIleHisGly 528

Search completed: September 16, 2003, 18:49:45
 Job time : 51 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:50:17 ; Search time 54 Seconds
(without alignments)
1480.662 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555

Perfect score: 2887

Sequence: 1 LFGSTPASVGRRLPKNPTG.....HTQSSVPLPLTATSMSSVGMA 537

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 556269 seqs, 148893369 residues

Total number of hits satisfying chosen parameters: 556269

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	517.5	17.9	421	9	US-09-420-785A-4
2	517.5	17.9	421	9	US-09-901-252-15
3	361	12.5	476	14	US-10-084-018-3
4	359	12.4	476	10	US-09-909-320-164
5	359	12.4	476	10	US-09-909-088B-164
6	359	12.4	476	10	US-09-905-291A-164
7	359	12.4	476	10	US-09-902-853-164
8	359	12.4	476	10	US-09-907-824-164
9	359	12.4	476	10	US-09-907-841-164
10	359	12.4	476	11	US-09-904-011-164
11	359	12.4	476	11	US-09-906-742-164
12	359	12.4	476	11	US-09-906-838-164
13	359	12.4	476	11	US-09-907-613-164
14	359	12.4	476	11	US-09-907-942-164
15	359	12.4	476	11	US-09-796-753-40

16	359	12.4	476	11	US-09-904-859-164	Sequence 164, App
17	359	12.4	476	11	US-09-909-204-164	Sequence 164, App
18	359	12.4	476	11	US-09-904-820-164	Sequence 164, App
19	359	12.4	476	11	US-09-904-786-164	Sequence 164, App
20	359	12.4	476	11	US-09-906-646-164	Sequence 164, App
21	359	12.4	476	11	US-09-906-700-164	Sequence 164, App
22	359	12.4	476	11	US-09-903-786-164	Sequence 164, App
23	359	12.4	476	11	US-09-902-903-164	Sequence 164, App
24	359	12.4	476	11	US-09-903-749A-164	Sequence 164, App
25	359	12.4	476	11	US-09-904-119-164	Sequence 164, App
26	359	12.4	476	11	US-09-904-956-164	Sequence 164, App
27	359	12.4	476	11	US-09-902-736-164	Sequence 164, App
28	359	12.4	476	11	US-09-907-794-164	Sequence 164, App
29	359	12.4	476	11	US-09-903-943-164	Sequence 164, App
30	359	12.4	476	11	US-09-904-463-164	Sequence 164, App
31	359	12.4	476	11	US-09-907-925-164	Sequence 164, App
32	359	12.4	476	11	US-09-902-692-164	Sequence 164, App
33	359	12.4	476	11	US-09-903-520-164	Sequence 164, App
34	359	12.4	476	11	US-09-905-056-164	Sequence 164, App
35	359	12.4	476	11	US-09-909-064-164	Sequence 164, App
36	359	12.4	476	11	US-09-904-553-164	Sequence 164, App
37	359	12.4	476	11	US-09-905-381-164	Sequence 164, App
38	359	12.4	476	11	US-09-905-088-164	Sequence 164, App
39	359	12.4	476	11	US-09-907-575-164	Sequence 164, App
40	359	12.4	476	11	US-09-905-075-164	Sequence 164, App
41	359	12.4	476	11	US-09-902-759-164	Sequence 164, App
42	359	12.4	476	11	US-09-902-634-164	Sequence 164, App
43	359	12.4	476	11	US-09-902-713-164	Sequence 164, App
44	359	12.4	476	11	US-09-907-979-164	Sequence 164, App
45	359	12.4	476	11	US-09-902-615-164	Sequence 164, App

ALIGNMENTS

RESULT 1

US-09-420-785A-4

; Sequence 4, Application US/09420785A

; Patent No. US20010010923A1

; GENERAL INFORMATION:

; APPLICANT: MOLESEN, UFFE

; APPLICANT: OLESEN, KJELD

; APPLICANT: STENNICKE, HENNING

; APPLICANT: BREDDAM, STEEN B.

; APPLICANT: BREDDAM, KLAUS

; TITLE OF INVENTION: MODIFIED CARBOXYPEPTIDASE

; FILE REFERENCE: 089187/0109

; CURRENT APPLICATION NUMBER: US/09/420,785A

; CURRENT FILING DATE: 1999-10-19

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 421

; TYPE: PPT

; ORGANISM: Saccharomyces cerevisiae

US-09-420-785A-4

Query Match 17.9%; Score 517.5; DB 9; Length 421;

Best Local Similarity 29.5%; Pred. No. 1.5e-40;

Matches 142; Conservative 70; Mismatches 185; Indels 85; Gaps 17;

QY 33 RYKEPGAEGVCEITPGVKSYSGYVDTSPF-SHTFFWFARHNPETAPITLWNGGSGD 91

DB 1 KIKDKPKILGI---DPNVTQYTYGLDVEDKHHFFWFESRNDPAKDPVILWNGGPGCS 57

QY 92 SLTGLFELGFCVHNSTFDYINPHSWNEVSNLFLSQPLGVGFYSDDTVGDSINPTGV 151

DB 58 SLTGLFELGFCVHNSTFDYINPHSWNEVSNLFLSQPLGVGFYSDDTVGDSINPTGV 104

QY 152 VENSPPAGVGPRPTIDATLIDTNTLAAEAWEILQGLSLGSLDSRVSOKDFSLWTES 211

DB 105 -----GSSG-----VSNTVAAGKDVYNFLFELFDQFPFVYVKNKGQ--DFHIAGES 146

QY 212 YGCHYPAFPHFYEQNERIANGSVNGVQLNFNSLGI--NGIIDEAIAQPIYFEFA-- 266
DB 147 YAGHYLPVFASEILSHKDR-----NFNLTSVLIGNGLDPLTOINYYEPMACGE 195
QY 267 -----VNNTYGIKAVNEVTYNYMKFANOMPNGCODLSTCKQNRRLALADYALCAEATNM 321
DB 196 GGEPSVLPSECSAMEDSL-----ERCLGLTESCYDSQ-----SWMSCVPATII 239
QY 322 CRNVEGPIYAFAGRGYVDIRHPYDD-----PTPPSYNKFSLAKDSVMDAIGVNNVITQS 376
DB 240 CNAQAAPYQOR-TGRNYYDIRKCEGNNLCYPTIQD-IDDYLNODYVKEAVGAEDVHYES 297
QY 377 -NNDVYAFQQTGFVMPNFIEDLEEILALPVRVSLIYGADYICNMFGGQAVSLAANY 435
DB 298 CNFIDNINFLPAGDMKP-YHTAVTDLLNQDLPILVYAGDKDFICNWLGNKAWTDVLPWK 356
QY 436 QAAQFRS---AGYTPLKVGVEYGETREYGNFSTRVYEAGHEVPYQPIASLQLFNRTI 492
DB 357 YDEEFASQKVRNWTASITDEVA-GEVKSYPKHTYLRVFNHGGMVFPDYPENALSMVNEWI 415
QY 493 FG 494
DB 416 HG 417

RESULT 2

US-09-901-252-15
; Sequence 15, Application US/09901252
; Patent No. US2002026658A1
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: Genes Encoding Sinapoylglucose:Malate Sinapoyltransferase and Met
; FILE REFERENCE: N1422-004
; CURRENT APPLICATION NUMBER: US/09/901,252
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent version 3.0
; SEQ ID NO 15
; LENGTH: 421
; TYPE: PRI
; ORGANISM: Saccharomyces
US-09-901-252-15

Query Match 17.9%; Score 517.5; DB 9; Length 421;
Best Local Similarity 29.5%; Pred. No. 1.5e-40;
Matches 142; Conservative 70; Mismatches 185; Indels 85; Gaps 17;

QY 33 RYKEPGAEGVCTTPGVKSYSGYVDTSP-SHTFFWFFEARHPETAPITLWNGGPGSD 91
DB 1 KIKDKPILGI---DPNVTQTYGLDVEDEKHHFFWTFESRNDPAKDPVILWNGGPGS 57
QY 92 SLIGLFEELGPCHVNSTFDYINPHSWNEVSNLFLSQPLQGVGSYSDIVDGSINPVTVG 151
DB 58 SLTGLFEELGPPSIGPDLKPIGNPYSWNSNATVIFLDQPVNNGFSYS----- 104
QY 152 VENSSTAGVGRYPTTDAITLDTNLAAEAWEILQGLSLDLSRVQSKDFSLWTES 211
DB 105 -----GSSG-----VSNTVAAGKDVYNFLELFDQFPYVKNKG--DFHAGES 146
QY 212 YGCHYPAFPHFYEQNERIANGSVNGVQLNFNSLGI--NGIIDEAIAQPIYFEFA-- 266
DB 147 YAGHYLPVFASEILSHKDR-----NFNLTSVLIGNGLDPLTOINYYEPMACGE 195
QY 267 -----VNNTYGIKAVNEVTYNYMKFANOMPNGCODLSTCKQNRRLALADYALCAEATNM 321
DB 196 GGEPSVLPSECSAMEDSL-----ERCLGLTESCYDSQ-----SWMSCVPATII 239
QY 322 CRNVEGPIYAFAGRGYVDIRHPYDD-----PTPPSYNKFSLAKDSVMDAIGVNNVITQS 376
DB 240 CNAQAAPYQOR-TGRNYYDIRKCEGNNLCYPTIQD-IDDYLNODYVKEAVGAEDVHYES 297

QY 377 -NNDVYAFQQTGFVMPNFIEDLEEILALPVRVSLIYGADYICNMFGGQAVSLAANY 435
DB 298 CNFIDNINFLPAGDMKP-YHTAVTDLLNQDLPILVYAGDKDFICNWLGNKAWTDVLPWK 356
QY 436 QAAQFRS---AGYTPLKVGVEYGETREYGNFSTRVYEAGHEVPYQPIASLQLFNRTI 492
DB 357 YDEEFASQKVRNWTASITDEVA-GEVKSYPKHTYLRVFNHGGMVFPDYPENALSMVNEWI 415
QY 493 FG 494
DB 416 HG 417

RESULT 3

US-10-084-018-3
; Sequence 3, Application US/10084018
; Publication No. US20020160499A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hawkins, Phillip R.
; Hillman, Jennifer L.
; Lal, Preeti
; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN SERINE
; CARBOXYPEPTIDASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,018
; FILING DATE: 25-Feb-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,689A
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/828,488
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0241 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 476 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MWLR3DT01
; CLONE: 566993
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-084-018-3

Query Match 12.5%; Score 361; DB 14; Length 476;
Best Local Similarity 26.7%; Pred. No. 1.5e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

QY 47 PG--VKYSYG--VDTSPSTHTFFWFFEARINPETAPITLWNGGPGSDSLIGLFEELGP 102
DB 67 PGLNMKSYAGPLTVNKTNSNLTFFWFFPAQIQPADPVVLQGGPGSSMXGLFVEHGP 126

Qy	103	CHVNSTFDYINPHSWNEVSNNLLFSLQPLGVGFYSYSDTVDGSI	NPVTGVVENSFAGVG	167
Dd	127	YWVTSMNTRLRDRDPWTTXSMXYIDNPVGTFSTDDTHG----		
Qy	163	RYPITADLTDTNLAAEAMIELLOFGLSPLDSRVSKDPSLWTESYGHHGPAPFN	222	
Dd	168	-----YAVEDDOVAROLYSALIOFF-----QIFPEYKNDFFYTGESYACKYPAIAH	215	
Qy	223	HFYEQNERIANGSVNGVLNFNSLGIIINGIDEAIIQAIPYEPFAVNNYGIKAVNETVYN	282	
Dd	216	LIHSLN-----PVREVKINLNGAIGDGYSPESIGCYAEF----	YQIGLLDEKKOKK 265	
Qy	283	YMKEANQMPNCODLLISTCKOIN---RYALADYAICABAIN--MCRDNVEG--PYATAFAG	335	
Dd	266	Y-FQKO-----CHCEIEHIRKONWFEAPEILDKILDGLTSDPSYFQNVTCGSNTYNFL-	318	
Qy	336	RGVDIRHPYDDTPPPSYNKFLAKDSYMDAIGNINYQTQSNDVY--YAFQOTGDFTVP	393	
Dd	319	-----RCTEPEDQLYYKFSLPEVRQAIHVGNQTFNDGTIVETKYLRDTSQSKP	369	
Qy	394	NFIEDLEIIALPVRSLSIYGDADI-----CNWFGGOAVSILAANYSAQA--	439	
Dd	370	WLTEIMNN-----YKVLIYNGQLDIIVAAAALTERSLMGMWDKGSQ-----	EYKKAEXV 418	
Qy	440	---FRS-----AGYPLKLVNGVEYGETREYGFNFSTRVYEACHEVPYPIASQLFNRTII	492	
Dd	419	WKLFKSUSEVAGY-----IROVGDFHQVIIRGGGHILFYDPOPLRAFDMINRFI	466	
Qy	493	F--GWD	496	
Dd	467	YKGWD	472	

RESULT 4

US-09-909-320-164
Sequence 164, Application 05/05909320
Patent No. US20020132240A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/909,320
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07

```

? PRIOR APPLICATION NUMBER: US 60/145,698
? PRIOR FILING DATE: 1999-07-26
? PRIOR APPLICATION NUMBER: US 60/146,222
? PRIOR FILING DATE: 1999-07-28
? PRIOR APPLICATION NUMBER: PCT/US99/20594
? PRIOR FILING DATE: 1999-09-08
? PRIOR APPLICATION NUMBER: PCT/US99/20944
? PRIOR FILING DATE: 1999-09-13
? PRIOR APPLICATION NUMBER: PCT/US99/21090
? PRIOR FILING DATE: 1999-09-15
? PRIOR APPLICATION NUMBER: PCT/US99/21547
? PRIOR FILING DATE: 1999-09-15
? PRIOR APPLICATION NUMBER: PCT/US99/23089
? PRIOR FILING DATE: 1999-10-05
? PRIOR APPLICATION NUMBER: PCT/US99/28214
? PRIOR FILING DATE: 1999-11-29
? PRIOR APPLICATION NUMBER: PCT/US99/28313
? PRIOR FILING DATE: 1999-11-30
? PRIOR APPLICATION NUMBER: PCT/US99/28564
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US99/28565
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US99/30095
? PRIOR FILING DATE: 1999-12-16
? PRIOR APPLICATION NUMBER: PCT/US99/30911
? PRIOR FILING DATE: 1999-12-20
? PRIOR APPLICATION NUMBER: PCT/US99/30999
? PRIOR FILING DATE: 1999-12-20
? PRIOR APPLICATION NUMBER: PCT/US00/00219
? PRIOR FILING DATE: 2000-01-05
? NUMBER OF SEQ ID NOS: 423
? SEQ ID NO 164
? LENGTH: 476
? TYPE: prt
? ORGANISM: Homo sapiens
? US-09-909-320-164

```

Query Match	12.4%	Score 359;	DB 10;	Length 476;
Best Local Similarity	26.7%	Pred. No. 2.3e-25;		
Matches	130;	Conservative 62;	Mismatches 178;	Indels 116; Gaps 21;
QY	47	PG--VKSYGY--VDTSFESHFWFFEARHNPTAPITLWLNGGPGSDSLIGLFEELGP	102	
Db	67	PGLNMKSYAGELVFNKTYNSLNFWFPPAQIQPEDAPWLWLQGGPGGSMFGLFVEHGP	126	
QY	103	CHVNSFTDDYINPHSNWEVSNLLFSLQPLGVGYSYSTVDGSINPVTGVVENSFAGVQG	162	
Db	127	YVVISNMTLKDREPPWTTIISMLIYDNPVGTGSFTDDTHG	167	
QY	163	RYPIIDATLDTTNLAAEAAWEILQGLPSLDSRVQSKDFSLWTESYGGHYGPAFFN	222	
Db	168	-----YAVNEDDVARDLYSALLQFF-----QIFPEYKKNDRYVYGESYAGYVPAIAH	215	
QY	223	HFEQNERTANGSVNGVOLNFSLIGTINGIIDBAIQAPYYPEFVAVNTYGIKAVNETVYN	282	
Db	216	LIHSLN-----PVREVKLNGLAIGDGYSDPESTIIGGYAER-----LYQIGLLDEKQKK	265	
QY	283	YMKFANQMPNGCODLISCTCKTN--RTALADYALCAEATN--MCRDNVEG--PYAFAG	335	
Db	266	Y--FQKQ-----CHECIEHRRKQNWFAFILDKLLEDGLTSDPSYEQNTTGGSNYYNFL-	318	
QY	336	RGVYIDIRHPYDDPTPPSYNKKFLAKDSVMDAIGVNIYNTQSNNDVY--YAFQQTGDFVWP	393	
Db	319	-----RCIPEPDOLYYVKFLSLPEVRCALHVG--NQTFNGDTIVEKYLREDTVQSVKPK	369	
QY	394	NFIEDLEELIALPVVRSLIYGDAHYI-----CNWFEGGQAVSLAANKYSAQAQ--	439	
Db	370	WLTEIMNN-----YKVLIIYNGQLDIIVAAALTELSLGMGMWKGSQ-----EYKRAEKV	418	
QY	440	---FRS-----AGYIPLKVNQVEGETREYGNFSFTRVYEAGHEVPYQPIASLQLFNRTI	492	
Db	419	WKIFXSDSEVAGY-----IRAGDEHQVILIRGGGHILPVDOELAFRDMFNRTI	466	

	APPLICANT:	Paoni, Nicholas F.	
	APPLICANT:	Roy, Margaret Ann	
	APPLICANT:	Stewart, Timothy A.	
	APPLICANT:	Tumas, Daniel	
	APPLICANT:	Williams, P. Mickey	
	APPLICANT:	Wood, William I.	
	TITLE OF INVENTION:	Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same	
	FILE REFERENCE:	10466-14	
	CURRENT APPLICATION NUMBER:	US/09/905/291A	
	PRIOR FILING DATE:	2001-07-12	
	PRIOR APPLICATION NUMBER:	PCT/US00/04414	
	PRIOR FILING DATE:	2000-02-22	
	PRIOR APPLICATION NUMBER:	US 60/143,048	
	PRIOR FILING DATE:	1999-07-07	
	PRIOR APPLICATION NUMBER:	US 60/145,698	
	PRIOR FILING DATE:	1999-07-26	
	PRIOR APPLICATION NUMBER:	US 60/146,222	
	PRIOR FILING DATE:	1999-07-28	
	PRIOR APPLICATION NUMBER:	PCT/US99/20594	
	PRIOR FILING DATE:	1999-09-08	
	PRIOR APPLICATION NUMBER:	PCT/US99/20944	
	PRIOR FILING DATE:	1999-09-13	
	PRIOR APPLICATION NUMBER:	PCT/US99/21090	
	PRIOR FILING DATE:	1999-09-15	
	PRIOR APPLICATION NUMBER:	PCT/US99/21547	
	PRIOR FILING DATE:	1999-09-15	
	PRIOR APPLICATION NUMBER:	PCT/US99/23089	
	PRIOR FILING DATE:	1999-09-15	
	PRIOR APPLICATION NUMBER:	PCT/US99/28214	
	PRIOR FILING DATE:	1999-11-29	
	PRIOR APPLICATION NUMBER:	PCT/US99/28313	
	PRIOR FILING DATE:	1999-11-30	
	PRIOR APPLICATION NUMBER:	PCT/US99/28564	
	PRIOR FILING DATE:	1999-12-02	
	PRIOR APPLICATION NUMBER:	PCT/US99/28565	
	PRIOR FILING DATE:	1999-12-02	
	PRIOR APPLICATION NUMBER:	PCT/US99/30095	
	PRIOR FILING DATE:	1999-12-16	
	PRIOR APPLICATION NUMBER:	PCT/US99/30911	
	PRIOR FILING DATE:	1999-12-20	
	PRIOR APPLICATION NUMBER:	PCT/US99/30999	
	PRIOR FILING DATE:	1999-12-20	
	PRIOR APPLICATION NUMBER:	PCT/US00/00219	
	PRIOR FILING DATE:	2000-01-05	
	NUMBER OF SEQ ID NOS:	423	
	SEQ ID NO 164		
	LENGTH:	476	
	TYPE:	PRT	
	ORGANISM:	Homo sapiens	
	US-09-905-291A-164		
	Query Match	12.4%; Score 359; DB 10;	
	Best Local Similarity	26.7%; Pred. No. 2.3e-25;	
	Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;		
QY	47 PG--VKSYSGV-VDTSPESHFFWFPEARINPTAITIWLNGPGSDSLIGLFEELGP	102	
Dd	: : : : : : :		
Dd	67 PLGNKSTAGLTWNKTYSNSLFMTFTPAQLQPEDADPVLWLGQGGGSMFGLVEHGP	126	
QY	103 CHVNSTFDYDYNPHSWNEWSNLLFLSPLGVGRSVTVDGSNPVTGVNVSFGAVOG	162	
Dd	: : : : : : : : : :		
Dd	127 YVTSNMTRDRDPFWITLSMLIDNPVGTFSTDDTHG-----	167	
QY	163 RPYTTDLTDNTNLAAFAWEIILOGLSLPSLDSSRVQSDFSLMTESYGGHYGAFTN	222	
Dd	: : : : : : : : : : : : : : : : : : : : :		
Dd	168 -----YAVEDDVARDLYASALIQF-----QIFPEYKNNDVFVTCESYAGKYVAIAH	215	
QY	223 HFYEONERIANGSNVOLFNISLIINGIIDAIQAIPYPFAVNNTYGIKAIVETVYN	282	
Dd	: : : : : : : : : : : : : : : : : : : : :		
Dd	216 LIHSLN-----PVREKNIINGIALGDGYSDPESIYGAAEF----LQIGILLDEKKK	265	
QY	283 YMKANOMPNCODLISTCKQTN---RTLADVALLCAEAATN---MCRDNVGE---PPYAFA	335	

Query Match	Score	DB ID	Length	Annotation
12.4*	359	476		SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Best 100	359	476		SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

Matches	130;	Conservative	62;	Mismatches	178;	Indels	116;	Gaps	21;
QY	47	PG--VKSYSGY--VDTSPSHTFTWFEEARHNPETAPITILWNGPGSDSLIGLFEELGP	102						
Db	67	GLNKKSYAGFTVKNYNTNSLNFEEFPAQIQEDAFVVLWLOGGPGGSMFGLFEVHGP	126						
QY	103	CHVNSTFDDYINPHSMKNEVSNLLFLSQPLGVGYSYDVTGVSINPVYIGVVENSSEFAGVQG	162						
Db	127	YVVTSMNWLDRDFPMTTWLSMLYIDNPVGTGTFSDDDTHG-----	157						
QY	163	RYFTIATLIDTNTLAAEAWEYLQGLFSLGSLDSRVQSKFSLMTESYGGHYGPAFFN	222						
Db	168	-----YAVNEDDVARDLYSALIQQF-----QIFPEYKNNDYVVTGESYAGKYVFAIAH	215						
QY	223	HFYEONERTANGSVNGVQLENFSLGTINGILDEAICQAPYPPEAVANTYGIKAVNETVYN	282						
Db	216	LTHSLN-----PVREVKNLNGIALGDSYSPESIIGYAF-----LQIGILLDEKQKK	285						
QY	283	YMKFANOMFENGQDILSTCKQTQN---RTALADYALCAEATN---MCRDNVEG--PYAFAG	335						

RESULT 8
US-09-907-824-154
; Sequence 164, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc

RESULT 8
US-09-907
; Sequenc
; Publica
; GENERA
; APPLIC
; APPLIC
; APPLIC
; APPLIC

```

QY 47 PG--VKSYSGY--VDISPSHTFFFEARHNPEAPITLWLNGSGSDSLIGLPEELGP 102
Db 67 PGLNKSAGFTVANKYNSLNLFEEFPAQIQPEDAPVVLWLGQGGSGSMFGLFVEHP 126
QY 103 CHVNSTFDDYINPHSWNEVSNLLFSLQPLGVGFSYSDIVDGSINPVIYGVNENSRAGVQG 162
Db 127 YVVTSMNLRDRDFPWTTLTSLMLYIDNPVGTGFSFTDDTHG----- 167
QY 163 RYPTIDATLIDTNLAEEAAWEILQGLSLGSLDSRVQSKDFSLWTESYGGHYGPAPFN 222
Db 168 -----YAVNEDDVARDLYSALIOFF-----QIFPEYKNNDFFYVIGESYAGKYVPAIAH 215
QY 223 HFYEONERLANGSVNGVOLNENSLGIINGIDEAIAQAPYYPFAVNNTYGIKAVNETVYN 282
Db 216 LIHSLN-----PVREVKINLNGIAIGDYSDPESIIIGGYAEF-----LYQIGLLDEKQKK 265
QY 283 YMKFANOMPNGQDLISCTKOTN---RTALADYALCAEATN---MCRDNVEG---PYAFAG 335
Db 266 Y--FQKQ-----CHECIEHQRKQNMFEAFIIDLKLDGLTSDPSYFQNVGTGCSNYNLF- 318
QY 336 RGVIDIRHPYDDPTPPSYNKKFLAKDSYMDAIGVNIYNTQSNNDVY--YAFQQTGDFWVP 393
Db 319 -----RCTEPEDQLYVVKFSLPEVROAIHVG-NQTFNDGTIIVKYLREDTIVOSVKP 369
QY 394 NFIEDLEILALPVRSVLIYGDADYI-----CNWFGQAVSLAANYSQAAQ-- 439
Db 370 WLTEIMNN-----YKVLINQGLDIIVAAALTERSLMGMDWKGSQ-----EYKKAQKV 418
QY 440 ---FRS-----AGYTPKLVNGVEYGETREYGNFSTRYVEAGHEVPYIQTASLOLFNRTI 492
Db 419 WKIFKSDSEVAGY-----IRQAGDFHQVIRGGGHILPYDQPLRAFDMINRFI 466
QY 493 F--GWD 496
Db 467 YGKGWD 472

```

RESULT 9

```

US-09-907-841-164
; Sequence 164, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jeannie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US/09/907,841
; GENERAL INFORMATION:

```

```

; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-841-164

```

```

Query Match 12.4%; Score 359; DB 10; Length 476;
Best Local Similarity 26.7%; Pred. No. 2,3e-25;
Matches 130; Conservative 62; Mismatches 176; Indels 116; Gaps 21;

```

```

QY 47 PG--VKSYSGY--VDISPSHTFFFEARHNPEAPITLWLNGSGSDSLIGLPEELGP 102
Db 67 PGLNKSAGFTVANKYNSLNLFEEFPAQIQPEDAPVVLWLGQGGSGSMFGLFVEHP 126
QY 103 CHVNSTFDDYINPHSWNEVSNLLFSLQPLGVGFSYSDIVDGSINPVIYGVNENSRAGVQG 162
Db 127 YVVTSMNLRDRDFPWTTLTSLMLYIDNPVGTGFSFTDDTHG----- 167
QY 163 RYPTIDATLIDTNLAEEAAWEILQGLSLGSLDSRVQSKDFSLWTESYGGHYGPAPFN 222
Db 168 -----YAVNEDDVARDLYSALIOFF-----QIFPEYKNNDFFYVIGESYAGKYVPAIAH 215
QY 223 HFYEONERLANGSVNGVOLNENSLGIINGIDEAIAQAPYYPFAVNNTYGIKAVNETVYN 282
Db 216 LIHSLN-----PVREVKINLNGIAIGDYSDPESIIIGGYAEF-----LYQIGLLDEKQKK 265
QY 283 YMKFANOMPNGQDLISCTKOTN---RTALADYALCAEATN---MCRDNVEG---PYAFAG 335
Db 266 Y--FQKQ-----CHECIEHQRKQNMFEAFIIDLKLDGLTSDPSYFQNVGTGCSNYNLF- 318
QY 336 RGVIDIRHPYDDPTPPSYNKKFLAKDSYMDAIGVNIYNTQSNNDVY--YAFQQTGDFWVP 393
Db 319 -----RCTEPEDQLYVVKFSLPEVROAIHVG-NQTFNDGTIIVKYLREDTIVOSVKP 369
QY 394 NFIEDLEILALPVRSVLIYGDADYI-----CNWFGQAVSLAANYSQAAQ-- 439
Db 370 WLTEIMNN-----YKVLINQGLDIIVAAALTERSLMGMDWKGSQ-----EYKKAQKV 418
QY 440 ---FRS-----AGYTPKLVNGVEYGETREYGNFSTRYVEAGHEVPYIQTASLOLFNRTI 492
Db 419 WKIFKSDSEVAGY-----IRQAGDFHQVIRGGGHILPYDQPLRAFDMINRFI 466
QY 493 F--GWD 496
Db 467 YGKGWD 472

```

RESULT 10

```

US-09-904-011-164
; Sequence 164, Application US/09904011
; Publication No. US2003003530A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hauspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavich, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-904-011-164

Query Match 12.4%; Score 359; DB 11; Length 476;
Best Local Similarity 26.7%; Pred No. 2,3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

QY 47 PG--VKSYSY--VDTSPESHIFFEAFRHNHPETAPITLWLNKGGPGSSSLIGFEELGP 102
Db 67 PGLNMSYAGFLTVNKTYSNLFNFWFPAQIQPEDAPVLWLGQGGSSMFGIFVHRGP 126
QY 103 CHVNSTFDDYINPHSNWVSNLLFLSOPLGVSYSYDVGSIINPVIGVYVENSFAGVQG 162
Db 127 YVYTSNMLRDRDPPTTILSMLYIDNPVGTGFSFTDDTHG----- 167
QY 163 RYPTIDATLIDTNLAARAAWEILQGLSLGSLDSRVQSKDSLTWSTESYGGHYGPAFFN 222
Db 168 -----YAVNEDDVARLDYSALIQFF-----QIFPEYKKNDFVYTGESAGKYVPAIAH 215
QY 223 HEYEONERIANGSVNGVQLNFNSLGIINGIIDEIAIQAPYYPEFAVNNTYGIKAVNETVYN 282
Db 216 LIHSLN-----PVREVKINLINGIAIGDYSDPESIIGGYAEF-----LYQIGLIDDEKQK 265
QY 283 YMKFANOMPGQDLISTCKQTN---RTALADYALCAETN---MCRDNVEG--PYXAFAG 335
Db 266 Y--PQKQ---CHECIEHIRKQNWFEAFELDKLIDGLTSDPSYFQNVGTGCSNYNLF- 318
QY 336 RGVDYIRHPYDDPTPPSYNKKFLAKDSVMDAIGVNIQSNNDVY--YAFQQTGDFVWP 393
Db 319 -----RCTEPEDQIYYKFLSLPEVRQAIHVG--NOTFNDGTIVEKYLREDIVQSVKP 369
QY 394 NFIEDLEILLALPVRSVLIYGDADYI-----CNWFGQAVSLAANYSOAAQ-- 439
Db 370 WLTEIMNN-----YKVLINQGLDIIVAAALTSRLMGMWKGSQ-----EYKKAERKV 418
QY 440 ---FRS-----AGYTLKVNVEYGETREYGNFSTRVVEAGHEVPYPIASLQLFNRTI 492
Db 419 WKIEKSDSEVAGY-----IROAGDFHQVIIRGGHILPYDQPLRAFADMINRFI 466
QY 493 F--QWD 496
Db 467 YGKGWD 472

RESULT 11
US-09-906-742-164
; Sequence 164, Application US/09906742
; Publication No. US20030023054A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hauspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavich, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
```



```
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-838-164

Query Match      12.4%; Score 359; DB 11; Length 476;
Best Local Similarity 26.7%; Pred. No. 2,3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

47 PG--VKSYSGY--VDTSPESTHTFFWFEARHNPETAPITLWLNKGGPGSDSLICGLFEELGP 102
   || :|||:| : : : :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
67 PGLNMKSYAGFLTVNKTYNSNLFWFEPQIQPEDAPVVLWLOGGPGGSSMFLGFVEHGP 126

103 CHVNSTFDDYINPHSNWNEVSNLLFSLQPLGVGSFSDIVDGSINPVTGVVENSFAVGQ 162
127 YVVTSMNLTDRDRFPWTTTSLMLYIDNPVGTGSFTDDTHG----- 167
163 RYPTIDATLDTINLAAEAAWEILOGLFSLGSLDSRVQSKDFSLWESYGGHYGAFFN 222
168 -----YAVNEDDVARDLYSALIQFE-----QIFPEYKKNDDYVVTGESYAGYVPAIAH 215
223 HFYEQNERIANGSVNGQLNFNSLGIITINGIDRAIOAPYPERAVNNYTIKAVNETVYN 282
216 LIHSLN-----PVREKVINLNGIAGDYSDPESIIGGYAEF-----LYQIGLLDEXQKK 265
283 YMKFANQMPNGCDLITCKQTN--RTALADYALCAEATN--MCRDNVVG--PYAFAG 335
266 Y--FQKO-----CHECIEHRIKQNNFEAFELDKLGDLSDPISFYONVIGCSNYNLF- 318
336 RGVDYDIRHPYDDPPPSYNNKFLAKDSVMDAIGVINYTOSNNVY--YAFQOQGFDFWVP 393
319 -----RCTEPEDQLYYVKFLSLPEVRQAIHVG-NQTFNDGTIVEKYLREDTVQSVKP 369
394 NFIEDLEELALPVRVSLIYGDADYI-----CNWFGGQAVSLAANYSOAAQ-- 439
370 WLTEIMNN-----YKVLIIYNGQLDIIYAAALTERSLMGMDWKGSQ-----EYKAEKKV 418
440 ---FRS---AGYTPLKNGVVEYGETREYGNFSTRVVEAGHEVPYPOPIASLQIFNRTI 492
419 WKIFKDSSEVAGY-----IRQAGDFHOVIINGCGHILPDYDQPLRAEDWINEFI 466
493 F--GWD 496
467 YGKGWD 472

RESULT 13
US-09-907-613-164
; Sequence 164, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fond, Sherman

; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 164
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-613-164

Query Match      12.4%; Score 359; DB 11; Length 476;
Best Local Similarity 26.7%; Pred. No. 2,3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

47 PG--VKSYSGY--VDTSPESTHTFFWFEARHNPETAPITLWLNKGGPGSDSLICGLFEELGP 102
   || :|||:| : : : :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
67 PGLNMKSYAGFLTVNKTYNSNLFWFEPQIQPEDAPVVLWLOGGPGGSSMFLGFVEHGP 126

103 CHVNSTFDDYINPHSNWNEVSNLLFSLQPLGVGSFSDIVDGSINPVTGVVENSFAVGQ 162
127 YVVTSMNLTDRDRFPWTTTSLMLYIDNPVGTGSFTDDTHG----- 167
163 RYPTIDATLDTINLAAEAAWEILOGLFSLGSLDSRVQSKDFSLWESYGGHYGAFFN 222
168 -----YAVNEDDVARDLYSALIQFE-----QIFPEYKKNDDYVVTGESYAGYVPAIAH 215
223 HFYEQNERIANGSVNGQLNFNSLGIITINGIDRAIOAPYPERAVNNYTIKAVNETVYN 282
216 LIHSLN-----PVREKVINLNGIAGDYSDPESIIGGYAEF-----LYQIGLLDEXQKK 265
283 YMKFANQMPNGCDLITCKQTN--RTALADYALCAEATN--MCRDNVVG--PYAFAG 335
266 Y--FQKO-----CHECIEHRIKQNNFEAFELDKLGDLSDPISFYONVIGCSNYNLF- 318
336 RGVDYDIRHPYDDPPPSYNNKFLAKDSVMDAIGVINYTOSNNVY--YAFQOQGFDFWVP 393
319 -----RCTEPEDQLYYVKFLSLPEVRQAIHVG-NQTFNDGTIVEKYLREDTVQSVKP 369
394 NFIEDLEELALPVRVSLIYGDADYI-----CNWFGGQAVSLAANYSOAAQ-- 439
370 WLTEIMNN-----YKVLIIYNGQLDIIYAAALTERSLMGMDWKGSQ-----EYKAEKKV 418
440 ---FRS---AGYTPLKNGVVEYGETREYGNFSTRVVEAGHEVPYPOPIASLQIFNRTI 492
419 WKIFKDSSEVAGY-----IRQAGDFHOVIINGCGHILPDYDQPLRAEDWINEFI 466
493 F--GWD 496
467 YGKGWD 472
```

```

Db 127 YVTSNMTLRDRDPFWTTLSMLYIDNPVGTGFSFTDDTHG----- 167
QY 163 RYPTIDATLIDTTNLAAEAAWEILQGLFSLDSRVQSKDFSLWTSYSGHYGPAPFN 222
Db 168 -----YAVNEDDVARLYSALIOFF-----QIFPEYKNDFFYVTGESYAGKYVPAIAH 215
QY 223 HFYEQNERIANGSVNGVQLNSGLIINGIIDEAIOAPYPEFVANNNTYGIKAVNETVYN 282
Db 216 LIHSLN-----PVREVKINLNGIAGDGYSDPESIIGGYAEF-----LYQIGLLDEKQKK 265
QY 283 YMKFANOMPNGCCDLISTCKQTN---RTALADYALCAEATN---MCRDNVEG---PYAFAG 335
Db 266 Y--FQKQ-----CHECHIRKQNWFEAFELDKLLDGLTSDPSYFQNVTCNSYNYFL- 318
QY 336 RGVYDIRHPYDDPTPPSYNKNFLAKDSYMDAIGNVINYTOSNNDVY--YAFQOTGDFVWP 393
Db 319 -----RCTEPEQIYYKFLSLPEVROAIHVG-NQTFNDGIVEXYLEDIVQSVKP 369
QY 394 NFIEDLEIILPVRVSLIYGADYI-----CNWFGQAVSLAANYSQAAQ-- 439
Db 370 WLTEIMNN-----YKVLIIYNGQLDIIVAAALTERSLMGMDWKGSQ-----EYKKAQKV 418
QY 440 ---FRS-----AGYTPLKVNQVEYGETREYGNFSTRVYEAGHEVPYQPIASLOLFNRTI 492
Db 419 WKIFKSDSEVAGY-----IROAGDFHQVILIRGGGHILPYDQPLRAFDMINRFI 466
QY 493 F--GWD 496
Db 467 YKKGWD 472

```

RESULT 14

```

US-09-907-942-164
; Sequence 164, Application US/09907942
; Publication No. US20030027146A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gottsien, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,942
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222

```

```

Query Match 12.4%, Score 359; DB 11; Length 476;
Best Local Similarity 26.7%, Pred. No. 2.3e-25;
Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;

QY 47 PG--VKSYGY--VDISPESHTEFFFEARINPETAPITWLNGPGSDSLIGLFEELGP 102
Db 67 PGLNKKSYAGFLTVKNTYNSLNFEEFPAQIQEDAPVVLWLOGPGSGSMFGLFVEHCP 126
QY 103 CHVNSTEDDYINPHSWNEVSNLLFLSOPLGVPFSYSDTVDSGINPVTGWENSSFAGVQG 162
Db 127 YVTSNMTLRDRDPFWTTLSMLYIDNPVGTGFSFTDDTHG----- 167
QY 163 RYPTIDATLIDTTNLAAEAAWEILQGLFSLDSRVQSKDFSLWTSYSGHYGPAPFN 222
Db 168 -----YAVNEDDVARLYSALIOFF-----QIFPEYKNDFFYVTGESYAGKYVPAIAH 215
QY 223 HFYEQNERIANGSVNGVQLNSGLIINGIIDEAIOAPYPEFVANNNTYGIKAVNETVYN 282
Db 216 LIHSLN-----PVREVKINLNGIAGDGYSDPESIIGGYAEF-----LYQIGLLDEKQKK 265
QY 283 YMKFANOMPNGCCDLISTCKQTN---RTALADYALCAEATN---MCRDNVEG---PYAFAG 335
Db 266 Y--FQKQ-----CHECHIRKQNWFEAFELDKLLDGLTSDPSYFQNVTCNSYNYFL- 318
QY 336 RGVYDIRHPYDDPTPPSYNKNFLAKDSYMDAIGNVINYTOSNNDVY--YAFQOTGDFVWP 393
Db 319 -----RCTEPEQIYYKFLSLPEVROAIHVG-NQTFNDGIVEXYLEDIVQSVKP 369
QY 394 NFIEDLEIILPVRVSLIYGADYI-----CNWFGQAVSLAANYSQAAQ-- 439
Db 370 WLTEIMNN-----YKVLIIYNGQLDIIVAAALTERSLMGMDWKGSQ-----EYKKAQKV 418
QY 440 ---FRS-----AGYTPLKVNQVEYGETREYGNFSTRVYEAGHEVPYQPIASLOLFNRTI 492
Db 419 WKIFKSDSEVAGY-----IROAGDFHQVILIRGGGHILPYDQPLRAFDMINRFI 466
QY 493 F--GWD 496
Db 467 YKKGWD 472

```


Db 383 TVDNFSLKVGAGHEVYQPTALQAFKQII-----QKK 418

RESULT 2

S51516
serine-type carboxypeptidase (EC 3.4.16.-) 2 precursor - Absidia zychar
C:Species: Absidia zychar
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Jun-2000
C:Accession: S51516; S78013; S78014
R:Lee, B.R.; Takeuchi, M.; Kobayashi, Y.
Curr. Genet. 27, 159-165, 1995
A:Title: Molecular cloning and sequence analysis of the sop2 gene encoding the serine ca
A:Reference number: S51516; MUID:95308538; PMID:7788719
A:Accession: S51516
A:Molecule type: DNA
A:Residues: 1-460 <LEE>
A:Cross-references: EMBL:D16519; NID:g556466; PIDN:BAA03966.1; PID:g995456
A:Experimental source: strain NRIC 1199
A:Accession: S78013
A:Molecule type: protein
A:Residues: 52-62;90-99;367-381 <LEB>
A:Accession: S78014
A:Molecule type: mRNA
A:Residues: 18-460 <LES>
C:Genetics:
A:Gene: spcZ
A:Introns: 64/3; 95/1; 107/2; 138/3; 234/2; 246/1; 270/1; 343/3; 353/2; 384/3; 437/3
C:Superfamily: serine carboxypeptidase
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-51/Domain: propeptide #status predicted <PRE>
F:52-460/Product: serine-type carboxypeptidase #status experimental <MAT>
F:40,46,119/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:186/Active site: Ser #status predicted

Query Match 25.9%; Score 746.5; DB 1; Length 460;
Best Local Similarity 35.8%; Pred. No. 3.5e-44;
Matches 175; Conservative 71; Mismatches 182; Indels 61; Gaps 14;
QY 6 PASVGRQLPKNTGVKTLTANNVIRYKEPAGVCEITPGVKSYGVDTSPESHTEF 65
Db 29 PMSHVLRQDNDT-----SSGNTLTQLKTSF---KLCD--PDVKQSYGLVDANDSHYF 78
QY 66 FWFEEARHNPEATITLWNGGSGSLIGLFEELGPGCHVNSTFDDYINPHSWNEVSNLL 125
Db 79 FWFEEKNDKPNPLIWLNGGPGCSLLIGLWEELGPGQNGS---ANPESWHHSNML 134
QY 126 FLSQPLGVGSYSVDVDSINPVTVGVVENSFACVQGRYPTIDATLIDTNLAEEANEI 185
Db 135 FFDQPDGVGTSY-----GKQ-----TVSTTEDAERAWTF 164
QY 186 LQGLSGIPSLDSRVQSKDPSLWTESYGGHYGPAFFNFHFEONERIANGVNGVQLNFS 245
Db 165 LQAFYETFP---QYSKLDVHYGESYGGHYIPGFASHVVDVMDNKKVSGEGKGVVPLKS 220
QY 246 LGIINGITDAIAQPYPEFVANNNTYGIKAVNETVYINMKFANQMPNGCQDLSTCKQTN 305
Db 221 IGVNGFLDAVIOYKSPKTKCHSTYPAVLSE---ECDKMQQIYENDCKPAAEQCAESD 277
QY 306 RTALADYALCAEATNMCRONVEGYPYAFAGVGVYDIHRPYDDPTPPSYNKFLLAKDSMD 365
Db 278 -----EDSDCVNANQCC-GQTEG-IYAQSGYSFYDIRQCGDD-TPHPFVDR-LNKASYIK 328
QY 366 AIGVNIQNTQSNNDVYAFQOTGDFVWPNFTEDLEEIALLPVRSYLIYGDADYICNNFEG 425
Db 329 EVGARGHFSMCSDSVGTAFAQTGDCA-RSYIPAVEKLLKEGIPVLIIYVGDADVLCNWCN 387
QY 426 QAVSLAANYSOAAQFRSAGYTPLVKNGVVEYGETREYGNFSFTRYEAGHEVYYPQIASL 485
Db 388 LDVADSLKWDGSAFASKLEAWKADGKEVGQFRSADKLTEVRVYEAGHEVPMYQPEAL 447
QY 486 QLFNKTIFG 494

Db 448 SMFQTWISG 456

RESULT 3

I43236
carboxypeptidase C (EC 3.4.16.5) precursor [validated] - fission yeast (Schizosacchar
N:Alternate names: carboxypeptidase Y
C:Species: Schizosaccharomyces pombe
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 08-Sep-2000
C:Accession: T43236; T37997
R:Tabuchi, M.; Iwaihara, O.; Ohtani, Y.; Onuchi, N.; Sakurai, J.; Morita, T.; Iwahara
J. Bacteriol. 179, 4178-4189, 1997
A:Title: Vacuolar protein sorting in fission yeast: cloning, biosynthesis, transport,
A:Reference number: Z22357; MUID:97352672; PMID:9209031
A:Accession: T43236
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1002 <TAB>
A:Cross-references: EMBL:D86560; NID:g3046860; PIDN:BAA25568.1; PID:g3046861
R:Oliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1997
A:Reference number: Z21760
A:Accession: T37997
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1002 <OLI>
A:Cross-references: EMBL:Z97209; PIDN:CABL0121.1; GSPDB:GN00066; SPDB:SPACL9G12.10c
A:Experimental source: strain 972h(-); cosmid c19G12
C:Comment: The mature carboxypeptidase exists as a heterodimer held together by a dis
C:Genetics:
A:Gene: cpy1; SPACL9G12.10c
A:Map position: 1
C:Superfamily: serine carboxypeptidase
C:Keywords: disulfide bond; glycoprotein; heterodimer; hydrolase; serine carboxypepti
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-521/Domain: propeptide #status predicted <PRO>
F:200-332/Region: 13-residue repeats (H-H-[EK]-B-G-E-H-M-P-P-[MF])
F:344-425/Region: 9-residue repeats (H-H-[RKQ]-G-G-[KE]-[DE]-K)
F:522-1002/Product: carboxypeptidase C #status experimental <MAT>
F:627-880/Disulfide bonds: #status predicted
F:659/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:715,921,978/Active site: Ser, Asp, His #status predicted

Query Match 22.6%; Score 652.5; DB 2; Length 1002;
Best Local Similarity 33.8%; Pred. No. 3.4e-37;
Matches 165; Conservative 66; Mismatches 174; Indels 83; Gaps 15;
QY 31 TIRYKEPAGVCEITPGVKSYGVDTSPESHTEFNFHFEARHNPEATITLWNGGPGS 90
Db 570 TLRVKDSKPSGLGDT--VKQYTGVLVDDEDRHLEFFWFESRNDPNDPVVLWNGGPGC 627
QY 91 DSLAGLFEELGPGCHVN-STFDDYINPHSWNEVSNLLFSLQPLGVGSYSVDVDSINPVT 149
Db 628 SSLAGLEMEIGLPGSSINIELTKPEYNPHSWNSNASVIFLDQPIINTGFSNGD--DSVLDITV 685
QY 150 GVWENSFAGVQGRYPTIDATLIDTNLAEEANEILOGELSGPLSLDSRVQSKDPSLWT 209
Db 686 -----AGKDVYAEALNLFKAFP-----QTAHLDFHAG 713
QY 210 ESYGCHYGPAPFNHFEONER-----LANG-SVNGVQLNFSLIGTINGLIDEAIAQAPYPE 264
Db 714 ESYAGHYIPOPAKEIMEHNQCANFVSGYEMEKQYINLKSVLINGLTDPLVQYFYEGK 773
QY 265 FAVNNYTG-----IKAVNETVYINMKFANQMPNGCQDLSTCKQTNRIALADYAL 314
Db 774 MACSPYGIINSQECEDRIIGAYDT-----CAKLITGYQGTGT-----PV 814
QY 315 CAEATNCRDNVEGYPYAFAGVGVYDIHRPYDDPTPPSY-----YNKFLAKDSVMDAIGV 369
Db 815 CIGASLCNNAMIGP-FTKTGLNIYDIREECDQHECHLYCPETGATESILNQEFVEALCV 873
QY 370 NINYTQSNNDVYAFQOTGDFVWPNFTEDLEEIL--ALPVRVSLTY-GDADYICNWFEGG 426

Db 874 EVDYKGCNEVNIQFLPKGDWMRKTRDDVIAILEAGLPV---LIYAGDADYICNMGNE 930
 QY 427 AVSLAANYSQAAQFRSAGYTPILKVNQVEYGETREYGNFSFTRVYAGHEVYQPIASLQ 486
 Db 931 AWTDALEWAGQEFYEALKPWSPNKZAGRGKSKNFYGLKLYEAGHVPFNQPEASLE 990
 QY 487 LENRTIFG 494
 Db 991 MLNSWIDG 998

RESULT 4

SA6008

Probable carboxypeptidase C (EC 3.4.16.5) YBR139w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR1015; hypothetical protein YBR139w
 C:Species: Saccharomyces cerevisiae
 C:Date: 26-Aug-1994 #sequence_revision 05-Jan-1996 #text_change 19-Apr-2002
 C:Accession: S46008; S46581
 R:Becam, A.M.; Herbert, C.J.; Nasr, F.; Slonimski, P.P.; Zagulski, M.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45995
 A:Accession: S46008
 A:Molecule type: DNA
 A:Residues: 1-508 <BEC>
 A:Cross-references: EMBL:Z36008; NID:g536435; PIDN:CAA85097.1; PID:g536436; GSPDB:GN0000
 A:Experimental source: strain S288C
 R:Becam, A.M.; Cullin, C.; Grzybowska, E.; Lacroute, F.; Nasr, F.; Ozier-Kalogeropoulos,
 Yeast 10(Suppl.A), S1-S11, 1994
 A:Title: The sequence of 29.7Kb from the right arm of chromosome II reveals 13 complete
 A:Reference number: S46569; MUID:94378717; PMID:8091856
 A:Accession: S46581
 A:Molecule type: DNA
 A:Residues: 1-508 <BE2>
 A:Cross-references: EMBL:X75891; NID:g496856; PIDN:CAA53497.1; PID:g496869
 A:Experimental source: strain S288C
 C:Genetics:

A:Cross-references: SGD:S0000343
 A:Map position: 2K
 A:Note: MIPS:YBR139w
 C:Superfamily: serine carboxypeptidase
 C:Keywords: hydrolase; serine carboxypeptidase; transmembrane protein
 F:6-22/Domain: transmembrane #status Predicted <TM>
 F:219,415,474/Active site: Ser, Asp, His #status predicted

Query Match 22.0%; Score 636; DB 1; Length 508;
 Best Local Similarity 32.6%; Pred. No. 1.9e-36;
 Matches 167; Conservative 70; Mismatches 197; Indels 78; Gaps 14;

QY 1 LPGSTPASVGRRL---PKNPICVKILTTANNVTIRYKEPGAEGVCETTPGVKSYGYV 56
 Db 43 LPQNTQOTLKLRNHHDDPLFTTIFSSVDTDYSLRLRTVDPSSKLG---DTVKQSGYM 98
 QY 57 DTSPESHTFWEFEARHNPETAPITLWLGPGSDSLIGLFEELGCHVNSTEDDYNPH 116
 Db 99 DYKDSKHFFYWFESRNDPNDPIILWLGPGCCSFTGLLFEELGPGSSICADMKPIHPY 158
 QY 117 SNEVSNLLFLSQPLGVGSYSDTVDSINPVTGVVENSFAGVQGRYPTIDATLIDTIN 176
 Db 159 SWNNASMIFFLEQLGVGFSYGD-----EKVSSYK 188
 QY 177 LAARAAWEILQGLSLPSLDSRVSKDPSLWTEYGGHYGPAFFNHFEQN-ERLANGS 235
 Db 189 LQKDAYIFLELFEAFPHL---RSNDPHIAGESYAGHYIPIQIAEIVVKNER----- 239
 QY 236 VNGVQLNFSNLGTINGLIDEALQAPVYPEFV-NTYTGKAVNEIVNTYMKFANQPNPC 294
 Db 240 ----TFNLTSVMIGNITDPLQADYEPMACGKGHYEVLSSCECKMSKAAGR----- 290
 QY 295 QDLSTCKQTNRTALADYAL--CAEATNCRONVECPYAFAGRGVYDIRHEDDP----- 348
 Db 291 -----CRLNLKLCYASKSSLPCIVATACDSALLEPYIN-TGLNVYDIRGPCEDNSTDG 343
 QY 349 ---TPPSYNKFLAKDSVDAIGVNI-NYQSNNDVYFAQQTGDFVWPNFTEDLEEILA 404

Db 344 MCYTGLRVYDQZVNFPEVQETLGSVDVHNSGCDNDVFTGLFTGDSKPFQYIAELLN 402
 QY 405 LPVRVSLYGDADYICNMFNGGGAVALSAAANSQAQFRSAGYTP--LKVNGVEYGETREY 462
 Db 403 HNPVLIYAGDKDYICNWLGNHANSNELEWINKRYQRMRLRPWWSKETGEELGOVKNYG 462
 QY 453 NESFTRVYAGHEVYQPIASLQLENRTIFG 494
 Db 463 PFTFLRIYDAGHMVPYDQPEASLEMVNSWISG 494

RESULT 5

JC7666

Serine-type carboxypeptidase homolog precursor - Emericella nidulans

N:Alternate names: carboxypeptidase Y homolog
 C:Species: Emericella nidulans, Aspergillus nidulans
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: JC7666
 R:Ohsumi, K.; Matsuda, Y.; Nakajima, H.; Kitamoto, K.
 Biosci. Biotechnol. Biochem. 65, 1175-1180, 2001
 A:Title: Cloning and characterization of the cpvA gene encoding intracellular carboxy
 A:Reference number: JC7666; MUID:21333188; PMID:11440134
 A:Accession: JC7666
 A:Molecule type: mRNA
 A:Residues: 1-552 <OHS>
 A:Cross-references: DDBJ:AB051820
 C:Comment: This protein is an intracellular carboxypeptidase, which is localized in v
 C:Genetics:
 A:Gene: cpvA
 A:Introns: 165/1
 F:1-15/Domain: signal sequence (or 1-18, or 1-19) #status predicted <SIG>
 F:16-131/Domain: pro-sequence, vacuolar targeting signal (or 19-131, or 20-131) #stat
 F:131-132/Region: cleavage site, by Arg-11e #status predicted

Query Match 20.8%; Score 601; DB 2; Length 552;
 Best Local Similarity 31.6%; Pred. No. 5.6e-34;
 Matches 154; Conservative 77; Mismatches 175; Indels 82; Gaps 19;

QY 27 ANNVTIRYKEPGAEGVCETTPGVKSYGYVDTSP-ESHTFWEFEARHNPETAPITLWLN 85
 Db 127 AYDLRIKTKDPSSLGI---DPVKQYTGYLDDNENKHLFWFESRNDPANDPVVLWLN 183
 QY 86 GPGSDSLIGLFEELGCHVNSTEDDYNPHSNVSNLLFLSQPLGVGSYSDTVDSGI 145
 Db 184 GPGCCSLTGLFELGPGSSIDENIKPVINPYANNSASVTFDQPVNNGYSYS----- 236
 QY 146 NPVTGVVENSFAGVQGRYPTIDAILLDTNLAARAAWEILQGLSLPSLDSRVQSKDF 205
 Db 237 -----GSTVSDIV-AAGKDYALLTLFFKQPF-----EYARODF 269
 QY 206 SLWTEYGGHYGPAFFNHFEQNRIANGSVNGVQLNFSNLGTINGLIDEALQAPYYPEF 265
 Db 270 HIAGESYAGHYIPVFTSEILSHQK-----NINLKSVLINGLTLGTLQVEYTRPM 320
 QY 266 AVNNTYTGKAVNETVINYMKFANQPNNG-----QDLSTCKQTNRTALADYALCAEATNMC 322
 Db 321 ACCEG-GYPAVLDE-----SSCRSDNALGRCSMTESCYNSESA-----WVCVPASIYC 369
 QY 323 RONVEGPYAFAGRGVYDIRHEDDPDT-----PPSYNKFELAKDSVDAIGVNI-YTQS 376
 Db 370 NNALLAPYOR-TQNYIDVRGKCEDESNLGYKMGYVSEYLNKPEVRAAAGAEVDGSDC 428
 QY 377 NNDVYFAQQTGDFVWPN--NFTEDLEEILALPVRSLLY-GDADYICNMFNGGGAVALSAA 433
 Db 429 NFDINNFLFHGDWMPYHRLVPGSILE--QIPV---LIYAGDADFTCNMLGNKAWTEALE 483
 QY 434 YSQAQFRSAGYTPILKV-----NGVEYGETREYGNFSFTRVYAGHEVYQPIASLQLF 488
 Db 484 WFGHKEFAAAPMEDLKIVDNEHTGKGIQIKTHGNTFMRLYGGHGHVWMDQPEASLEFF 543
 QY 489 NRTIFG-W 495
 ||| :| :| :|

326 Db ERCLGLIESYDSQ-----SVMSCVPATLYCNAQLAPYQR-TGRNVYDIRKDCGGNLC 379

348 QY -PTPSPYXNKLAKSDVMDAIGVNLNYTQS-NNDVYFAQQTGDFWPNFIEDLEIILAL 405

380 Db YPTLQD-IDDYLNQDYKEAVGAEDVHYESCNDLNRNLFAGDMKPK-YHTAVTDLLNQ 437

406 QY PVRVSLIYGDAIYICNFGGQAVSLAANSYAAQAFRS---AGYTPLKNGVVEYGTREVG 462

438 Db DLPILVYAGDKDFCTCNMLGNKAWTDVLPKWKYDEEASQKVRNWTASIIDEVA-GEVSKYK 496

463 QY NFSTRVRYEAGHEVPYQPTIASLOLFNRTIFG 494

497 Db HFTYLRVFNCGHMVFPDVPENALSMVNEWIHG 528

RESULT 9

T03607

probable carboxypeptidase C (EC 3.4.16.5) cbp31 - rice

N:Alternate names: serine-type carboxypeptidase homolog

C:Species: Oryza sativa (rice)

C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 21-Jul-2000

C:Accession: T03607

R:Washio, K.; Ishikawa, K.

Plant Physiol. 105, 1275-1280, 1994

A:Title: Organ-specific and hormone-dependent expression of genes for serine carboxypeptidase in rice

A:Reference number: Z14975; MUID:95062718; PMID:7972496

A:Accession: T03607

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-429 <WAS>

A:Cross-references: EMBL:D17587; NID:g409581; PIDN:BAA04511.1; PID:g409582

A:Experimental source: cv. Yukihiikari

C:Genetics:

A:Gene: cbp31

C:Superfamily: serine carboxypeptidase

C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase

F:6,414,417/Binding site: carbohydrtate (Asn) (covalent) #status predicted

F:148,336,393/Active site: Ser, Asp, His #status predicted

Query Match 16.4%; Score 473; DB 1; Length 429;

Best Local Similarity 28.6%; Pred. No. 3e-25;

Matches 130; Conservative 77; Mismatches 178; Indels 70; Gaps 16;

QY 52 YGY--VDTSPESHTFFWFEARHP-ETAPITLWNGPGSDSLIGLEELGCPCHVNST 108

Db 17 HAGYRLPNDHARLRYFFESRSGEDDPVYILWLTGGCSEALFYENGPFHFDN 76

QY 109 FDDYINPHSWNEVSNLLFTSQPLGVGFSYSDTVGDSINPVGVNENSSFAGVGQRYPTID 168

Db 77 MSLVWDFGQDQESNLIYYDQPTGTGFSYSSNPRDTRHDEAG-VSNDLYA----- 125

QY 169 ATLIDTNLAAEAAWELLOGELGSLDLSRVSQSKDFSLWTSYGGHYGAFFNHHYEQN 228

Db 126 -----FLOAFTEHPNF-----AKNDFYITGESYAGHYIPAFASRYVYKGN 165

QY 229 ERIANGSVNGVQLNLSGLIINGIIDEALQAPYYPFAVNNVTYGIKAVNETVYNYMKEAN 288

Db 166 K-----NSEGIHNLKGFAIGNGLIDPAIQKAYTDYSLD-----MGLIYKSPNRI--N 213

QY 289 QMPNCGQDLISTCKOTNR-TALADYALCAEATNMCNDNVEGGYAFAGRGVYDIRHP--- 344

Db 214 KIVPTCELAIKLCGTSGITSLGAYVWC---NLICSSIE---TIIGKKNYVDIRKPCVG 266

QY 345 ---YDDPTPPSYNNFLAKDSYMDAIGV-NINYTSNNDVYVAFQGTGDFVWPNFTE-DL 399

Db 267 SLCYD-----LGNMEFLQKSVRESLGVGDIQFVSCSPTYQAML-----LDWMRNLVGI 318

QY 400 EEILALPYRVSLIYGDAIYICNFGGQAVSLAANSYAAQAFRSAGYTPKLVNGVEYGETR 459

Db 319 PELLENDIKVLIYAGEYDILCNWLGNSRWNSWENSGKEAFVSSSEEPPTVDGKEAGILK 378

QY 460 EYGNFSRTRVYEAAGHEVPYQPTIASLOLFNRTIFG 494

Db 379 SYGSLFLKLVHDAGHVPMDQPKVALEMLRWTSQ 413

RESULT 10
S22530
carboxypeptidase C (EC 3.4.16.5) precursor - rice
N:Alternate names: carboxypeptidase III
C:Species: Oryza sativa (rice)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Nov-1999
C:Accession: S22530
R:Washio, K.; Ishikawa, K.
Plant Mol. Biol. 19, 631-640, 1992
A:Title: Structure and expression during the germination of rice seeds of the gene for a
F:22-73/Domain: signal sequence #status predicted <PRO>
F:74-484/Product: carboxypeptidase C #status predicted <MAT>
F:485-499/Domain: carboxyl-terminal propeptide #status predicted <CPRO>
F:144/Binding site: carboxylate (Asp) (covalent) #status predicted
F:216,404,461/Active site: Ser, Asp, His #status predicted

Query Match 15.7%; Score 452; DB 1; Length 500;
Best Local Similarity 26.4%; Pred. No. 1.1e-23;
Matches 128; Conservative 81; Mismatches 214; Indels 62; Gaps 14;
QY 14 LPK--NPTGVKTLTANNVIRKEPCAGVCTTGVKSYSGY--VDTSPESHFTWFF 69
DB 47 LPKAGTGTAGDVPSPVAPGELLRRVTLPLGPGVGLGHAGYRPLNTHDARMTFLF 106
QY 70 EARHNPETAPITLWNGPGSDSLGLFEELGPGHVNSTFDDYINPHSNVEVSNLLFLSQ 129
DB 107 ESRGKKED-PVWILTGPGCSSELAVFYENGPTTISNNKSLANKEGWDFTISNIFVDQ 165
QY 130 PLGVGFYSYDVTGGINPVTGVVNSFAGVQGRYPIDATLIDTTNLAFAAWEILQCF 189
DB 166 PTGTGFGYSDDRDTRHDTGV-----SNDLYSFLQVF 198
QY 190 LSGPLSDRSVQKDFSLWTFESYGHYGPAPFNHFYENONRIANGSVNGVQNLNENSLGII 249
DB 199 FKKHPEF-----AKNDFITGESYAGHYIPAFASRVHQNK--AN---EGIHNLKGFALG 249
QY 250 NGIIDEATQAPYEPFAVNTYGIKAVNETVYNTMKFANQMPGQDLISTCKQTNRTAL 309
DB 250 NGLTDPALQYKAYTDYALD-----MNLKKSVDRI---NKEIPCEFAIKLGGTNGK--- 299
QY 310 ADYALCAEATNMCRDNVEGPPYAPAGRVYDRIHPYDPTPPSYN--KFLAKDSYMDAI 367
DB 300 ---ASCAAYMVCSIFSSIMKLVGTKNYDVRKECEGLKCYDFSNLEKFTGDKAVKEAL 356
QY 368 GV-NINYTOSNNDVYAFQOQGDFFWPNFIE-DLEEIILALPVRVSLIYGDADYICNWRFG 425
DB 357 GVGDLFFVSCSTTVQAMLD-----WMENLEVGIPALLEDGINVLIIYAGEYDLICNWLGN 412
QY 426 OAVSLAANYSOAOFPSAGYTLKYNVEYGETREYGNFSTFRVYEAAGHEVYYPPIASL 485
DB 413 SRWHISMWMSGKDKDFVSSHESFPVVDGAEAGVLKSHGSLPLFKVNHAGHVMVMDQPKASL 472
QY 486 QLFNR 490
DB 473 EMLRR 477

RESULT 11
T48977
carboxypeptidase-like protein F14D17.80 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 24-Oct-2000
C:Accession: T48977
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25008
A:Accession: T48977
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-510 <OR>
A:Cross-references: EMBL:AL353992; GSPDB:GN00061; ATSP:F14D17.80
A:Experimental source: cultivar Columbia; BAC clone F14D17
C:Genetics:
A:Gene: ATSP:F14D17.80
A:Map position: 3
A:Introns: 115/2; 171/3; 259/3; 310/1; 341/3; 428/1; 472/3
A:Superfamily: serine carboxypeptidase
F:231,421,478/Active site: Ser, Asp, His #status predicted

Query Match 15.2%; Score 439; DB 2; Length 510;
Best Local Similarity 26.6%; Pred. No. 8.8e-23;
Matches 123; Conservative 81; Mismatches 189; Indels 69; Gaps 14;
QY 46 TPGVKSY---SGY--VDTSPESHFTWFFEARHNPETAPITLWNGPGSDSLGLFEEL 100
DB 93 SPSVQDFGHGAGYKLPNSKAARMFYFFESRTN-KADPVVILWLTGGPGCSSELALFVN 151
QY 101 GPCVNSTFDDYINPHSNVEVSNLLFLSQPLGVGFYSYDVTGGINPVTGVVNSFAGV 160
DB 152 GPTVSNNSSLSWNEFGWDKASNLIIYDQPVGTGFSYTSQSDRLRHDDEGV----- 202
QY 161 QGRYPTIDATLIDTTNLAFAAWEILQGLSLGSLDSRVQSKDFSLWTFESYGHYGPAP 220
DB 203 -----SNDLYFLQAFKHEP-----QFVKNDFYITGESYAGHYIPAL 240
QY 221 FNHFYENONRIANGSVNGVQNLNENSLGIINGIDEAIOAPYYPFAVNTYGIKAVNETV 280
DB 241 ASRVHRGNK-----NKEGTHNLKGFALGNLTNPETIOYGAAYADYALDMLNLTQSDHNL 295
QY 281 YNYMKFANQMPGQDLISTCKQTNRTALADYA---LCAEATNMCRDNVEGPPYAFAGR- 336
DB 296 NRY-----YATQQSTKECSADGGEGDACCASYTVC--NNIFOKIMDIAGNV 340
QY 337 GYDTRHPYDPTPPSYN--KFLAKDSYMDAIGV-NINYTOSNNDVYAFQOQGDFFWVP 393
DB 341 NYDVRKQCEGLCVDFSNMENFLNQKSVKALGVGLDIEFVSCSTAVTEAMDMD---WM 396
QY 394 NFIE-DLEEIILALPVRVSLIYGDADYICNWRFGQAVSLAANYSOAOFPSAGYTLKYN 452
DB 397 RNLEVGIPALLEDGINVLIIYAGEYDLICNWLGNKSWHMEWMSGKQEFVAAATVEFHV 456
QY 453 VEYGETREYGNFSTFRVYEAAGHEVYYPPIASLQLFNRTIFG 494
DB 457 KEAGLMKNYGSITFLKLVHDAGHVPMDQPKAALOMLQNMWQ 498

RESULT 12
A35275
carboxypeptidase C (EC 3.4.16.5) - barley
N:Alternate names: carboxypeptidase III
C:Species: Hordeum vulgare (barley)
C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 24-Nov-1999
C:Accession: A35275
R:Sorensen, S.B.; Svendsen, I.; Bredam, K.
Carlsberg Res. Commun. 54, 193-202, 1989
A:Title: Primary structure of carboxypeptidase III from malted barley.
A:Reference number: MUID:90315015; PMID:2639682
A:Accession: A35275
A:Molecule type: protein
A:Residues: 1-411 <SOR>
A:Note: 185-Val was also found
C:Superfamily: serine carboxypeptidase
C:Keywords: blocked amino end; glycoprotein; hydrolase; serine carboxypeptidase

	Matches	150;	Conservative	85;	Mismatches	202;	Indels	173;	Gaps	26;
QY	4	STPASVGRRLPKNPTGVKTLTANNVTIRYKPGAEGVCEITPGV	-----KSYSGYVD	57						
Db	1521	STPAGI-----DVTQPLTNVISQINCT-----TQTDRIINLPGLPADMLFKQYSFELD	1570							
QY	58	TPSESHTFWFPEARHNPTAIIITLWNGPGSDSLIGLFEELGPGCHVSNFTDDYT-NPH	116							
Db	1571	GLSGHKHVYLVSENNPSTDPLLWNGPGSSLMGLFEENGPRFVSKDSQTLSENFY	1630							
QY	117	SNWSEYNLLFLSOPLCVGSYSDTVDGSIINPVTGVVENSFAGVQGRYPTIDATLIDTIN	176							
Db	1631	SNWKFANVLYLSPTGVGYSYA-----YNTNINQYDDVT-	1664							
QY	177	LAABAEWELLQGLSLGSLDSRVQSKDFSLMTESYGGHYGPAPFNFHYEQNERIANGSV	236							
Db	1665	-TAQENYAALKSFQAQY----QYTSDFEYTGESYAGVYLPGL-----SALLVQGIK	1712							
QY	237	NG-VOLNFNLSLGLINGIDE-----AIOAPYPERAVNNTYGIKAVNETVYN-----	282							
Db	1713	SGDININYGVSIGNVIDDKTDMNSQLHYQY-----HG--GISATFYATALOLCC	1762							
QY	283	-----YMKFANOMPNG-----QQD-LISTCKQTNRITALADYALCABATNKCERD	324							
Db	1763	SGDEFKCFRSDRMTNFNNISIPWCDLSPCYDFVWATGAQLLLTAEPDPYNNYQOCWTI--	1819							
QY	325	NVEGPPYAYAGRCYVDIRHPYDDPTPPSYNKK-----	FL 358							
Db	1820	-----PYNDTTPRTPYGETWTGINTYESSDFLNGPYCYDSDAMEGYL	1860							
QY	359	AKDSVMDALGV--NINYTQSNNDVYAFQGTGFVWPNFTEDELEETLA---LPVRVSLIY	413							
Db	1861	NRPVYRKALNIPDSVPYWAANNININAYNQVDSDIVPN---LQILMANAPAFKMLIYS	1916							
QY	414	GDAITYCNWFQGGQ-----AVSLAANTYSQAA---QFRSAGVYTPLVKNGVYGETREYG	462							
Db	1917	GDAITMVNNLGAEIFTANNFAALGLTSSPRAQWTFQYNST-FQP-TVAGYQYQSYTSNAI	1974							
QY	463	NFSFTRVYBAGHEVPYQYPIASLOLNRRTIFGWDIAEGOKKIWPSYKTNCTATATHTQSS	522							
Db	1975	NIDVLTVKSGSGHEVPLDRFQQAALQM-----IYNFVKSGRGINTPPLDLSNFTTTITTTSTTIT	2030							
QY	523	VP--LPTATS	530							
Db	2031	TPQGTGPTVTA	2040							

RESULT 15
T49079
serine-type carboxypeptidase like protein - Arabidopsis thaliana
N:Alternate names: protein F4F15.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T49079
R:Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25015
A:Accession: T49079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <ALC>
A:Cross-references: EMBL:AL049711; GSPDB:GNC0061; ATSP:F4F15.110
A:Experimental source: cultivar Columbia; BAC clone F4F15
C:Genetics:
A:Gene: ATSP:F4F15.110
A:Map position: 3
A:Introns: 116/1; 151/1
C:Superfamily: serine carboxypeptidase

Query Match 13.3%; Score 383; DB 2; Length 482;
Best Local Similarity 26.2%; Pred. NO. 6.3e-19;
Matches 135; Conservative 68; Mismatches 202; Indels 110; Gaps 18;

```
Search completed: September 16, 2003, 22:52:32
Job time : 40 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:35:07 ; Search time 33 seconds
(without alignments)
765.253 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555
Perfect score: 2887
Sequence: 1 LPGSTPASVGRRLPKNPTG.....HTQSSVPLPTATSMSSVGM 537

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	894	31.0	423	1	CPSL_PENJA
2	652.5	22.6	1002	1	CBPY_SCHPO
3	636	22.0	508	1	YBY9_YEAST
4	599	20.7	542	1	CBPY_CANAL
5	594.5	20.6	523	1	CBPY_PICPA
6	527.5	18.3	532	1	CBPY_YEAST
7	474	16.4	516	1	CBPX_ARATH
8	473	16.4	429	1	CBPX_ORYSA
9	452	15.7	500	1	CBP3_ORYSA
10	434	15.0	508	1	CBP3_HORVU
11	413.5	14.3	500	1	CBP3_WHEAT
12	359	12.4	476	1	CPVL_HUMAN
13	353.5	12.2	436	1	CP22_HORVU
14	352.5	12.2	470	1	YSS2_CAEEL
15	351	12.2	482	1	NF31_NAEFO
16	349	12.1	474	1	PRTP_MOUSE
17	344.5	11.9	516	1	YUW5_CAEEL
18	332	11.5	469	1	CP23_HORVU
19	328.5	11.4	480	1	PRTP_HUMAN
20	326.5	11.3	454	1	YUA6_CAEEL
21	321.5	11.1	729	1	KEX1_YEAST
22	318.5	11.0	476	1	CBP2_HORVU
23	316	10.9	574	1	YPP3_CAEEL
24	312.5	10.8	471	1	VCP_AEDAE
25	312	10.8	510	1	CBP1_ORYSA
26	305.5	10.6	423	1	CBP2_WHEAT
27	297.5	10.3	505	1	YXD2_CAEEL
28	295.5	10.2	286	1	CBPX_PEA
29	280	9.7	499	1	CBP1_HORVU
30	269.5	9.3	523	1	PEPS_ASPSA
31	255	8.8	452	1	RISC_MOUSE
32	248	8.6	452	1	RISC_RAT
33	237	8.2	452	1	RISC_HUMAN

34 221.5 7.7 531 1 PEPF_ASPNG
35 219.5 7.6 507 1 SXA2_SCHPO
36 167.5 5.8 324 1 CP21_HORVU
37 147 5.1 366 1 HNLS_SORBI
38 120.5 4.2 1061 1 OAR_MYXXA
39 114.5 4.0 717 1 AMYM_BACST
40 113.5 3.9 1287 1 RP01_FOWPV
41 112 3.9 376 1 OE56_NPVAC
42 112 3.9 787 1 AGL2_BACTQ
43 111 3.8 696 1 OXLA_NEUCR
44 110.5 3.8 1196 1 AMYB_PAEPO
45 109.5 3.8 481 1 PRTB_ERWCH

ALIGNMENTS

RESULT 1
CPSL_PENJA
ID CPSL_PENJA STANDARD; PRI; 423 AA.
AC P34946;
DI 01-FEB-1994 (Rel. 28, Created)
DI 01-FEB-1994 (Rel. 28, Last sequence update)
DI 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase S1 (EC 3.4.16.6).
OS Penicillium janthinellum (Penicillium vitale).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5079;
RN [1]
RP SEQUENCE.
RX MEDLINE=94039747; PubMed=6224168;
RA Svendsen L., Hofmann T., Endrizzi J., Remington S.J., Breddam K.;
RI "The primary structure of carboxypeptidase S1 from Penicillium
janthinellum";
RL FEBS Lett. 333:39-43(1993).
CC -!- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine
or lysine residue.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
DR PIR; S38953; S38953.
DR HSSP; P00729; LYSC.
DR MEROPS; S10.008; -.
DR InterPro; IPR000379; Ser_estrs_site.
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR PRODOM; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein.
FT ACT_SITE 143 143 POTENTIAL.
FT ACT_SITE 340 340 BY SIMILARITY.
FT ACT_SITE 397 397 BY SIMILARITY.
FT BINDING 343 343 SUBSTRATE.
FT BINDING 398 398 SUBSTRATE.
FT DISULFID 8 68
FT DISULFID 55 300
FT DISULFID 223 246
FT DISULFID 230 239
FT CARBOHYD 200 200 N-LINKED (GLCNAC...).
SQ SEQUENCE 423 AA; 46496 MW; 55F3333608B12729 CRC64;

Query Match 31.0%; Score 894; DB 1; Length 423;
Best Local Similarity 40.1%; Pred. No. 1.5e-55;

Matches 186; Conservative 66; Mismatches 160; Indels 52; Gaps 8;

QY 41 GVCEITPGVKSYSYVDTSPESTFFWFTEARINPETAPILNLNGSPGSDSLGLFEEL 100

Db 6 GICETTPGVNQYSGLSVGSNNWFWFEARNPQAPLAANFNGSGCGSSMIGLFEEN 65

QY 101 GPCH-VNSIFDDYINPHSWNEVSNLFLSPLGVGFSYSDTVGINSFVTVGVVSSSFG 159

Db 66 GPCHFVNGSTPSLNENSWNNYNIIDPIGVGFSYG----- 104

QY 160 VQGRYPTIDATLDTNLAARAEWELQGLSLGSRVQSKDFSLWTSYSGGHGPA 219
 Db 105 -----TDDTV-----STVTAPYVWNLQAFYAPRPEYER----DRAIFTESYGGHGP 151
 QY 220 FNFHFEONERKANGSVQVQLNFSNLGIINGIIDEAIOAPYYPPEFANNVNYGKAVNET 279
 Db 152 FASYIFEQNAAIKAGSVTGQNVIVAGVNGWIDSTIQEKAVIDFSYNNYSQOIIDSSI 211
 QY 280 VYNYMFANPMNGCQDLISTCKQTNKRTALADYALCAEATNMCRDNVGEGYAFAGGVY 339
 Db 212 RDSLLDAYN---NQCLPALQCCSQSGST-----SDCTNADSVQYQNIIEGPISSGDPDVY 263
 QY 340 DIRHPYDDPPEPSYINKFLAKDSVMDAIGVNNYQTSNNDDVYFAQQTQGFVWPNFTEDL 399
 Db 264 DIREPSNDYPPKTYSTYLSIDPTVVKALGARTNQECPNKPYNKEASTGDNPRSFSTL 322
 QY 400 EEILALFVRVSLYIGDADYICNFMFGQAVSLAANYSOAQAQFRSAGYTPPLKVNVEYGETR 459
 Db 323 SSVQSGINVLWAGDADWICNLGNVEVANAVDFPGNAQFSALDLPATYVNGVEKGQFK 382
 QY 460 EYGNFSFTRYEAGHEVPYQPTASLQFLNRTIFGWDIAEGOKK 503
 Db 383 TVDNFSFLKYGAGHEVPYQPTALQAFKQII-----QKK 418

RESULT 2
 CBPY_SCHPO STANDARD; PRT: 1002 AA.
 ID CBPY_SCHPO
 AC 013849; 014366;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase Y precursor (EC 3.4.16.5) (CYP).
 GN CPY1 OR CPY1L OR SPAC19G12.10C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID:4896;
 RN [1]
 RP SEQUENCE FROM N.A., SUBUNIT, SUBCELLULAR LOCATION, AND MUTAGENESIS OF
 RP CYS-627.
 RX MEDLINE=97352672; PubMed=9209031;
 RA Tabuchi M., Iwahara O., Ohtani Y., Ohuchi N., Sakurai J.-I.,
 RA Morita T., Iwahara S., Takegawa K.;
 RT "Vacuolar protein sorting in fission yeast: cloning, biosynthesis,
 RT transport, and processing of carboxypeptidase Y from
 RT Schizosaccharomyces pombe.";
 RL J. Bacteriol. 179:4179-4189(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Welltjens I., Volckaert E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer C., Moestl D., Hillbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 Shpakowski G.V., Usery D., Bartell B.G., Nurse P.;
 "The genome sequence of Schizosaccharomyces pombe.";
 Nature 415:871-880(2002).
 RL -!- FUNCTION: Involved in degradation of small peptides. Digests
 preferentially peptides containing an aliphatic or hydrophobic
 residue in p1' position, as well as methionine, leucine or
 phenylalanine in p1 position of ester substrate.
 CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 broad specificity.
 CC -!- SUBUNIT: Heterodimer of two subunits of 32 kDa and 19 kDa derived
 from the precursor protein and linked by a disulfide bond.
 CC -!- SUBCELLULAR LOCATION: Lysosome-like vacuoles.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D86560; BAA25568.1; -;
 CC EMBL: Z97209; CAB10121.1; -;
 CC PIR: T43236; T43236.
 CC HSSP: P00729; 1XSC.
 CC MEROPS: S10.001; -;
 CC GeneDB_Spombe: SPAC19G12.10C; -;
 CC InterPro: IPR000379; Ser_estr_site.
 CC InterPro: IPR001563; Serine_carbpept.
 CC Pfam: PF00450; serine_carbpept; 1.
 CC PRINTS: PR00724; CRBOXYPTASEC.
 CC PRODOM: P0001189; Serine_carbpept; 1.
 CC PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
 CC PROSITE: PS00560; CARBOXYPEPT_SER_HIS; FALSE_NEG.
 CC Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Repeat; Signal.
 FT SIGNAL 1 18
 FT PROPEP 19 521
 FT CHAIN 522 1002
 FT DOMAIN 208 341
 FT DOMAIN 225 341
 FT REPEAT 225 237
 FT REPEAT 238 250
 FT REPEAT 251 263
 FT REPEAT 264 276
 FT REPEAT 277 289
 FT REPEAT 290 302
 FT REPEAT 303 315
 FT REPEAT 316 328
 FT REPEAT 329 341
 FT DOMAIN 361 423
 FT REPEAT 361 369
 FT REPEAT 370 378
 FT REPEAT 379 387
 FT REPEAT 388 396
 FT REPEAT 397 405
 FT REPEAT 406 414
 FT REPEAT 415 423
 FT ACT_SITE 715 715
 FT ACT_SITE 921 921
 FT ACT_SITE 978 978
 FT BINDING 924 924
 FT BINDING 979 979
 FT DISULFID 627 860
 FT DISULFID 776 789
 FT DISULFID 799 822
 FT DISULFID 806 815

RX	MEDLINE=95042830; PubMed=7954890;
RA	Nasr F., Becam A.-M., Grzybowska E., Zagulski M., Sionimski P.P.,
RA	Herbert C.J.;
RT	"An analysis of the sequence of part of the right arm of chromosome
RT	II of <i>S. cerevisiae</i> reveals new genes encoding an amino-acid permease
RL	and a carboxypeptidase.";
RL	Curr. Genet. 26:1-7(1994).
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collabor-
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstat-
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no
CC	modified and this statement is not removed. Usage by and for commec-
CC	entities requires a license agreement (see http://www.isb-sib.ch/anno
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X75891; CAA53497.1; -;
DR	EMBL; Z36008; CAA85097.1; -;
DR	PIR; S46008; S46008.
DR	HSSP; P00729; LQPY.
DR	MEROPS; S10.UPW; -.
DR	SGD; S0000343; YBR139W
DR	InterPro; IPR000379; Ser_estr_site.
DR	InterPro; IPR001563; Serine_carbpept.
DR	Pfam; PF00450; serine_carbpept; 1.
DR	PRINTS; PR00724; CRBOXYPASEC.
DR	ProDom; PD001189; Serine_carbpept; 1.
DR	PROSITE; PS00131; CARBOXYPEPT_SER_HIS; 1.
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW	Hypothetical protein; Hydrolase; Carboxypeptidase.
FT	ACT_SITE 219 219 BY SIMILARITY.
FT	ACT_SITE 415 415 BY SIMILARITY.
FT	ACT_SITE 474 474 BY SIMILARITY.
SEQ	SEQUENCE 508 AA; 57639 MW; AAB2806CBEE2EDE1 CRC64;
Query Match	22.0%; Score 636; DB 1; Length 508;
Best local similarity	32.6%; Pred. No. 2.5e-37;
Matches 167; Conservative 70; Mismatches 197; Indels 78; Gaps	
QY	1 LPGSTASVGRQL-----PKNPTGVKLTITANNVIRYKEPGAOCVCTTPGVKSYSGY
DB	43 LPQNTQTLKLORLNHDDLFTEFTSSVDYDSLRLTVDPFSKLG----DTVKQWSGM
QY	57 DTSPESTTFWFEEARHNPTAPIIWLNGPGSDSLGLEELGCPCHVNSTFDYINPH
DB	99 DYKDSKHFFYWFESNDPANDPIIWLNGPGCSFTGLLEFGPSIGADMKPIHPNY
QY	117 SNEVSNLLFLSQPLGVGSFYSDTYVGSINPVTGVVENSFAGVGRYPTIDAFLDITN
DB	159 SWNNASMTIFLEQLGVGFSGD-----EKVSIIK
QY	177 LAEAAWEITLOGLSLPSLSRVOSKFSLWTESYGHYGPAPFNHFYEQN--BRIANGS
DB	189 LAGKDAIYPLELFFFAFPHL----RSNDFHLAGESYAGHYLPQIAHEITVWKNER----
QY	236 VNGVLNFNSLGIINGIDEAQAQYPEEAV--NNTYGIKAVNETVYNMKFANQMPCGC
DB	240 ----INFNLTSVMIGNITDPLQADYYEPMACGGKGHYHVLSSECEKMSKAAGR-----
QY	295 QDLISTCKOTNRATADYAL--CAEATNMCRDNVEGPPYAFAGRGVYDIRHPYDDP----
DB	291 -----CRRNLKCVASKSSLPCIVATAICDSALLEPIN-TGLNVYDTRGPCEDNSTDG
QY	349 ---IPPSYINKELAKSDYMDAIGVNI-NYTOSNNDVVYAFOQTGFVPNFIEDLEILA
DB	344 MCYTLGRYVDQYMNPFEVOETIGSDVHNSGCCDNDVFGLFTGDGSKP-FQQYIAELLN
QY	405 LPVRYSLIYGDADYICNNGFGQAVSLAANSQAQFRSAGYTP--LKVNGVYGETIREY
DB	403 HNPVLLIYAGDKDYICTNWLGHWASNELEINKERYQRMLRPWVSKETGEELGQVKNYG
Y	463 NFSFTFRYEAEGHEVPYQPIASLOLFNHTIEG 494

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87131100; PubMed=3028649;
RA Valls L.A., Hunter C.P., Rothman J.H., Stevens T.H.;
RI "Protein sorting in yeast: the localization determinant of yeast
RI vacuolar carboxypeptidase Y resides in the propeptide.";
RL Cell 48:887-897(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Dedman K.C., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Connor R., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Jagels K., Iye G., Moute S., Odell C., Pearson D., Rajadream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RI The nucleotide sequence of Saccharomyces cerevisiae chromosome
RI XIII.";
RT Nature 387:90-93(1997).
RL [3]
RP SEQUENCE OF 112-532.
RA Svendsen I., Martin B.M., Viswanatha T., Johansen J.T.;
RI "Amino acid sequence of carboxypeptidase Y. II. Peptides from
RI enzymatic cleavages.";
RL Carlsberg Res. Commun. 47:15-27(1982).
RN [4]
RP REVISIONS, AND ACTIVE SITE SER-257.
RA Breddam K., Svendsen I.;
RI "Identification of methionyl and cysteinyl residues in the substrate
RI binding site of carboxypeptidase Y.";
RL Carlsberg Res. Commun. 49:639-645(1984).
RN [5]
RP ACTIVE SITE HIS-508.
RX MEDLINE=90345013; PubMed=2639680;
RA Bech L.M., Breddam K.;
RI "Inactivation of carboxypeptidase Y by mutational removal of the
RI putative essential histidyl residue.";
RL Carlsberg Res. Commun. 54:165-171(1989).
RN [6]
RP MUTAGENESIS.
RX MEDLINE=94114555; PubMed=7904479;
RA Mortensen U.H., Remington S.J., Breddam K.;
RI "Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen
RI bond network stabilizes the transition state by interaction with the
RI C-terminal carboxylate group of the substrate.";
RL Biochemistry 33:508-517(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=95244421; PubMed=7727362;
RA Endrizzi J.A., Breddam K., Remington S.J.;
RI "2.8-A structure of yeast serine carboxypeptidase.";
RL Biochemistry 33:11106-11120(1994).
CC [1]- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS
CC PREFERENTIALLY PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC
CC RESIDUE IN P1' POSITION, AS WELL AS METHIONINE, LEUCINE OR
CC PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.
CC [2]- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC [3]- ENZYME REGULATION: INHIBITED BY 2PCk.
CC [4]- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
CC [5]- PTM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZYMOGEN AND
CC IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO
CC A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE
CC GOLGI, EXTENSION OF ITS CORE OLIGOSACCHARIDES LEADS TO THE
CC GOLGI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM
CC SECRETORY PROTEINS AT OR BEYOND A LATE GOLGI COMPARTMENT AND IS
CC SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR AND IS
CC ENDOSOME-LIKE COMPARTMENT. UPON ARRIVAL IN THE VACUOLE, THE
CC N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMICALLY
CC ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).

Db 326 ERLGLIESCYDSQ-----SVWSCVPEATYCNNAQLAPYOR-TGRNVYDIRKDCGGNLC 379

QY 348 -PTPPSYNFKIAKSDVMDAIGVINYNTOS-NDVYIAFOQTGDFWPNFIEDLEIAL 405

Db 380 YTLQD-IDYLNQDYVKGAEVHYHSCNFIDNRNLFAGDWKPK-YHTAVTDILNQ 437

QY 406 PYRSLIYGDADYICNWFQGVASIAANYSSQAQFRS---AGYTPLKVGVEYGETREYG 462

Db 438 DLPILVYAGDKDFICNMLGNKAWTDVLPWKYDEEFASQVRNWTASITIDEVA-GEVYSYK 496

QY 463 NFSFTRVYAGHEVPYQYPIASLQLENRIFG 494

Db 497 HFTYLVFNGGHMVPPDPVPENALSMVNEWIHG 528

RESULT 7

CBPX ARATH STANDARD; PRT; 516 AA.

AC P32826; Q42107; Q9CAE5;

DT 01-OCT-1993 (Rel. 27, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Serine carboxypeptidase precursor (EC 3.4.16.-)

GN AT3G10410 or F13M14.32.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bradley D.;

RL Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016720; PubMed=11130713;

RA Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,

RA Farmanbaf M., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,

RA Delsen M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,

RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,

RA Winkler P., Cattolico L., Weissenbach J., Saurh W., Quettier F.,

RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,

RA Wurbmann E., Drzonek H., Erfle H., Jordan N., Bangert S.,

RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,

RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,

RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,

RA Reichelt J., Scharfe M., Schoen O., Bagues M., Terol J., Climent J.,

RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,

RA Cooke R., Laudie M., Berger-Lilauro C., Purnelle B., Masuy D.,

RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,

RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,

RA Mayer K.F.X., Kaul S., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,

RA Rooney T., Rizzo M., Waits C., Koo H.L., Tallon L.J., Jenkins J.,

RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,

RA Pai G., Militscher J., Sellers P., Gill J.E., Feldblum T.V.,

RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,

RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,

RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,

RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,

RA Watanabe A., Yamada M., Yasuda M., Tabata S.;

RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana".

RL Nature 408:820-822(2000).

RP [3]

RP SEQUENCE OF 252-372 AND 455-516 FROM N.A.

RC STRAIN=cv. Columbia; TISSUE=Seedling;

RC Hofte H.;

RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a broad specificity.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M81130; AAB04606.1; -

DR EMBL; AC011560; AAG51389.1; -

DR EMBL; Z25955; CAA81115.1; -

DR EMBL; Z36528; CAA81299.1; -

DR HSSP; P00729; IYSC.

DR MEROPS; S10.009; -

DR InterPro; IPR000379; Ser_estrs_site.

DR InterPro; IPR001563; Serine_carbpept.

DR Pfam; PF00450; serine_carbpept; 1.

DR PRINTS; PR00724; CRBOXYPTASEC.

DR ProDom; PD001189; Serine_carbpept; 1.

DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.

DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.

KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.

FT SIGNAL 1 22 POTENTIAL.

FT PROPEP 23 82 POTENTIAL.

FT CHAIN 83 516 SERINE CARBOXYPEPTIDASE.

FT ACT_SITE 229 229 BY SIMILARITY.

FT ACT_SITE 417 417 BY SIMILARITY.

FT ACT_SITE 474 474 BY SIMILARITY.

FT BINDING 420 420 SUBSTRATE (BY SIMILARITY).

FT CARBOHYD 157 157 N-LINKED (GLCNAC...) (POTENTIAL).

FT DISULFID 139 379 BY SIMILARITY.

FT DISULFID 300 322 BY SIMILARITY.

FT DISULFID 307 315 BY SIMILARITY.

FT CONFLICT 515 516 QM -> ADVTSSFALHNKROQIIKQVSN (IN REF. 1).

SQ SEQUENCE 516 AA; 57301 MW; 177C778DF657A1C1 CRC64;

Query Match 16.4%; Score 474; DB 1; Length 516;

Best Local Similarity 30.1%; Pred. No. 6.1e-26;

Matches 139; Conservative 71; Mismatches 172; Indels 80; Gaps 18;

QY 45 TTPGVKYSQY-VDTSPSHFFWFEARHNDETAPITLWNGGPGSLSLGLPEELGP 102

Db 93 TVDDIGHAGYKLPKSRGASMEYFFESH-NKKDAPVVIWLGPGSCSELAFFVENG 151

QY 103 CHVNSTFDDYINPHSNWNSNLLFLSQPLGVGESYDITVDGSIINPVTVGVVENSFAGVQG 162

Db 152 FKITSNNKSLAWNEYQWDQVSNLLYVDPVGTGFSYITDKSDIRHDETVG----- 200

QY 163 RYPTIDATLIDTNNLAAEAWEILQGLSLGSLDSRVOSKDFSLWTSYGGHYGPAFFN 222

Db 201 -----SNDLYDFLQAFFAHPKL-----AKNDFYITGESYAGHYIPAFAS 240

QY 223 HFEYONERTANGSVNGVQLNFSNLGTLINGLIDEAIOAPYYPFAVNTYGIKAVNTYVN 282

Db 241 RVHKGNNK--AN---EGVHNLKGFALGNLTDPALQYPAIPDYALE--MGLITQKE---- 289

QY 283 YMKFANMPNGCQDLTSTCKQTNRTA-LADYALCAEATNMCNKNVGGPYAFAGRGVYDI 341

Db 290 HDRLEKIVPL-CELSIKLCTGDTTSCIASYVNCNLSFSCVMSHAGGVNV-----YDI 341

QY 342 RHP-----YDDTTPSYNNKFLAKDSVMDAIGV-NINYTSQNDNDVYAFQQTGDFWPN 394

Db 342 RKKCVGSLCYDF-----SNMEKFLNQSVRKSLGVGDIDFVSCSTSVYQA-----MLVD 390

QY 395 FIEDLEEILALPVRYS-----LIY-GDADYICNWFQGVASIAANYSSQAQFRSAGYTPL 448

Db 391 WMKNLE--VGIPITLLEDGIGSLVYAGEYDLICNWLGNRSWVNAWMSGKTFNCAKEVPF 448

QY 449 KVGVEYGETREYGNFSTRVYEAGHEVPYQYPIASLQLENR 490


```

FT PROPEP      22 73      BY SIMILARITY.
FT CHAIN       74 484
FT PROPEP      485 500      BY SIMILARITY.
FT ACT_SITE    216 216      BY SIMILARITY.
FT ACT_SITE    404 404      BY SIMILARITY.
FT ACT_SITE    461 461      BY SIMILARITY.
FT BINDING     407 407      SUBSTRATE (BY SIMILARITY).
FT CARBOHYD     144 144      N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE    500 AA; 55446 MW; AE455E2780147D88 CRC64;

Query Match      15.7%; Score 452; DB 1; Length 500;
Best Local Similarity 26.4%; Pred. No. 2.le-24;
Matches 128; Conservative 81; Mismatches 214; Indels 62; Gaps 14;

Qy 14 LPK--NPTGVKLTANNVIRYKEPGAECETTPGVKSYGY--VDTSPESHTFFWF 69
Dy 47 LPKEAGTGDVPSVAPGELLERRVTLPGLPQGVGLGHAGYRFLNTHDARMFYLE 106
Qy 70 EARHNPETAPITLWLNGSGSDSLGLFELGPGCHVNSTDDYINPHSWNEVSNLLFLSQ 129
Dy 107 ESRGKKED-PVVIWLTGGPGCSSELAIFYENGPTTISNNMSLANKFGWDTISNLIIFVDQ 165
Qy 130 PLGVGFSYSDTVGSGINPVTGVVNSFSFAGVQGYRPTIDATFLIDTTLNLAEEAAWEILQGF 189
Dy 166 PTGTFGFSYSSDDRDTRHDETV-----SNDIYSLQVF 198
Qy 190 LSGPLSDSRVQSKDFSLWTSYCHYGPAFFNHFYEQNERIANGSVNGVLNFSGLII 249
Dy 199 FKKHPEF-----AKNDFITGESYAGHYIPAPASRVHOGNK--AN---EGIHINKLGEAIG 249
Qy 250 NGIIDEATQAPYPEFAVNNYIGKAVNETVYNTWKTFANOMPNSGCDLISCKQTNRIAL 309
Dy 250 NGLTDPALQKAYTDYALD-----NMLIKSYDRI---NKFIPCEFAIKLCGGNGK--- 299
Qy 310 ADYALCAEATNMCRDNVEGPPYAFAGRVGYDIRHYDDPTPPSYN--KFLAKDSYMDAI 367
Dy 300 ---ASMAAYVWCNSIFSSIMKLGVTKNYIDVRKECEGKLCYDFSNLEKFFGDKAVKEAI 356
Qy 368 GV-NINTQSNNDVYPAQQTGDFWPNFIE-DLEEILALPVRSIYGDADYICNTPGG 425
Dy 357 GVGLEFVSCSTTVYQAMLTD---WNRNLEVGPALLEEDGINVLIVAGEVDLNCNLGN 412
Qy 426 QAVSLAANYCAAQFRAGYTPLVKNGVEYGETREYGNFSFTRVVEAGHEVPYTCIASL 485
Dy 413 SRWVHSWESGQKDFVSSHESFFVVDGAEAGVLKSHGPLSLFKVHNAGHVPMDQPKASL 472
Qy 486 QLFNR 490
Dy 473 EMLRR 477

RESULT 10
CBP3_HORVU STANDARD; PRT; 508 AA.
ID CBP3_HORVU
AC P21529;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-MIII).
GN CBP3 OR CXP;3.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Himalaya; TISSUE=Aleurone;
RA Kocher A., Lok F., Cameron-Mills V., von Wettstein D.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 81-491.
RC STRAIN=cv. Gula;

```

```

RX MEDLINE=90315015; PubMed=2639682;
RA Soerensen S.B., Svendsen I., Breddam K.;
RT "Primary structure of carboxypeptidase III from malted barley.";
RL Carlsberg Res. Commun. 54:193-202(1989).
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- ENZYME REGULATION: INHIBITED BY MERCURIC IONS.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOSPERM.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED MAINLY IN THE ALEURONE AND, TO A
CC LESSER EXTENT IN THE EMBRYO, THROUGHOUT THE 5-DAYS GERMINATION
CC PERIOD EXCLUSIVELY, WITH A MAXIMAL LEVEL AT 3 DAYS. ALSO FOUND IN
CC THE ROOTS AND SHOOTS OF THE GROWING SEEDLING.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Y09604; CAA70817.1; -.
HSP: P00729; ICPY.
MEROPS: S10.009; -.
InterPro: IPR000379; Ser_estrs_site.
InterPro: IPR001563; Serine_carbpept.
Pfam: PF00450; serine_carbpept; 1.
PRINTS: PR00724; CRBOXPTASEC.
ProDom: P001189; Serine_carbpept; 1.
PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 19
FT PROPEP 20 80
FT CHAIN 81 491 SERINE CARBOXYPEPTIDASE III.
FT PROPEP 492 508
FT MOD_RES 81 81 BLOCKED.
FT ACT_SITE 223 223 BY SIMILARITY.
FT ACT_SITE 411 411 BY SIMILARITY.
FT ACT_SITE 468 468 BY SIMILARITY.
FT BINDING 414 414 SUBSTRATE.
FT CARBOHYD 151 151 N-LINKED (GLCNAC... ).
FT VARIANT 265 265 Q -> V (POLYMORPHISM).
SQ SEQUENCE 508 AA; 56362 MW; 70C6751D78D40AB6 CRC64;

Query Match      15.0%; Score 434; DB 1; Length 508;
Best Local Similarity 26.5%; Pred. No. 3.9e-23;
Matches 132; Conservative 77; Mismatches 221; Indels 68; Gaps 15;

Qy 1 LPGSTPASVGRRLPKNETGVKLTANNVIRYKEPG-AEGVCTTPGVKSYGY--VD 57
Dy 47 LPKDSSSSSGRHG-ARVGEDEDVAPGOLLERRVTLPGLPGEVAD----LGHAGYRFLP 101
Qy 58 TSPESHTFFWFPEARHNPETAPITLWLNGSGSDSLGLFELGPGCHVNSTEDDYINPHS 117
Dy 102 NTHDARMFIFFESRGKKED-PVVIWLTGGPGCSSELAIFYENGFTTANNMSLVNKEG 160
Qy 118 WNEVSNLLFLSQPLGVGFSYSDTVGSGINPVTGVVNSFSFAGVQGYRPTIDATFLIDTTL 177
Dy 161 WDKISNIIFVDQPTGTGFSYSSDDRDTRHDETV----- 194
Qy 178 AEAWEILQGLSLGSLDSRVQSKDFSLWTSYCHYGPAFFNHFYEQNERIANGSVN 237
Dy 195 -SNDIYDFLQVFFKKHPEF-----IKNDFITGESYAGHYIPAPASRVHOGNKK-----NE 244
Qy 238 GVQLNFSGLIINGIIDEATQAPYPEFAVNNYIGKAVNETVYNTWKTFANOMPNSGCDL 297
Dy 245 GTHINKGFAICNGLTDPALQYKAYTDYALENMLNLCQADYERI-----NKFIPCEFA 297
Qy 298 ISTCKQTNK-TALADYALCAEATNMCNVEGPPYAFAGRVGYDIRHYDDPTPPSYN- 355

```

```

Db 298 IKLCGTNGKASCMAYVNCVTFNSIMKLVGKNY-----YDVRKECEGKLCYDFSNL 350
QY 356 -KFLAKDSVMDAIGV-NTNYTQSNNDVYAFQQTGDFVWPNFIE-DLEEIIALPVKVSILI 412
Db 351 EKFFGDKAVRQIGVGDIEFVSCSTSVQAMLT-----WMRNLEVGIPALLEDGINVLIY 406
QY 413 YGDADYICNWRGGQAVSLAANYSQAAQFRSAGYPLKYNVNGVEYGTREYGNFSTRVYDA 472
Db 407 AGEYDLICNWLGNRWHSMGWSGQKDFAKTAESSFLVDDAQAGVLSHGALSLFKVHNA 466
QY 473 GHEVPYQPIASLQLFNR 490
Db 467 GHMVPMDQPKAALMLRR 484

RESULT 11
CBP3_WHEAT
ID CBP3_WHEAT STANDARD; PRT; 500 AA.
AC P11515;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine carboxypeptidase III precursor (EC 3.4.16.5) (CP-WIII).
GN CBP3.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=86007602; PubMed=2820978;
RA Baulcombe D.C., Barker R.F., Jarvis M.G.;
RT "A gibberellin responsive wheat gene has homology to yeast
RT carboxypeptidase Y";
RL J. Biol. Chem. 262:13726-13735(1987).
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- SUBUNIT: Monomer (Probable).
CC -!- INDUCTION: By gibberellic acid (GA).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J02817; AAA34273.1; -.
DR PIR: A29412; A29412.
DR HSP: P00729; ICPY.
DR MEROPS: S10.009; -.
DR InterPro: IPR000379; Ser_estrs_site.
DR InterPro: IPR001563; Serine_carpept.
DR Pfam: PF00450; serine_carpept; 1.
DR PRINTS: PR00724; CBPOXITASEC.
DR ProDom: PD001189; Serine_carpept; 1.
DR PROSITE: PS00131; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
DR Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 21
FT PROPEP 22 73
FT CHAIN 74 484
FT PROPEP 485 500
FT ACT_SITE 216 216 BY SIMILARITY.
FT ACT_SITE 404 404 BY SIMILARITY.
FT ACT_SITE 461 461 BY SIMILARITY.
FT BINDING 407 407 SUBSTRATE (BY SIMILARITY).
FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 500 AA; 55334 MW; H2ACE10EF8484CDA CRC64;

```

```

Query Match 14.3%; Score 413.5; DB 1; Length 500;
Best Local Similarity 25.7%; Pred. No. 1e-21;
Matches 122; Conservative 75; Mismatches 201; Indels 77; Gaps 15;

QY 38 GAEGVCE-----TTPGVKS-----YSGY--VDTSPESTHFWFEARHNPEAPI 80
Db 58 GAEDVAPGOLLERRVTLPGLPBGVGLGHHAGYKRLPNTHDARMEYFFFEESGKED-PV 116
QY 81 TILWNGGPGSDSLIGLFEELGCHVNSTFDDYINPHSNNEVSNLFLSLQPLGVGSYSST 140
Db 117 VTLWGGGSCSELAVFYENGFTTANNMSLVNKFQWDKISNLIIVDPATGTGYSYSD 176
QY 141 VDGSTNPVTGVVENSFAGVQGRYPTIDATLIDTTLNLAEEAAWEILQGLFLSLPLSDSRV 200
Db 177 DRDTRHDEAGV-----SNDLYDFLQVFFKKHPEF----- 205
QY 201 QSKDFSLWTSYGHGYPAPFNHFEONEIRANGSVNGVOLNPNLSGIINGLIDRAIQAP 260
Db 206 VKNDFITGESYAGHYIFAFASRVHGNKK-----NEGTHLNKGAIGNGLTDFAIQYK 260
QY 261 YYPEFAVNTYGIKAVNETVYVNMKFANOMPNGCODLSTCKOTNR-TALADYALCAEAT 319
Db 261 AYTDYALD-----MNLIQADYDRI---NKFIPPCEFAIKLCGTGDKASCMAYVNCNSTF 313
QY 320 NMCNDNVRGPPYAFAGRGVYDIRHPYDDPTPEPSYTN--KFLAKDSVMDAIGV-NINYTQS 376
Db 314 NSIMKLVGKNY-----YDVRKECEGKLCYDFSNLEKFFGDKAVRQIGVGDIEFVSC 366
QY 377 NNDVYAFQQTGDFVWPNFIE-DLEEIIALPVRSLLIYGDADYICNWRGGQAVSLAANY 435
Db 367 STSVYQAMLT-----WMRNLEVGIPALLEDGINVLIYAGEYDLICNLSNRSWHSMEWS 422
QY 436 QAAQFRSAGYPLKYNVNGVEYGTREYGNFSTRVYAGHEVYQPIASLQLFNR 490
Db 423 GQKDFAKTAESSFLVDDAQAGVLSHGALSLFKVHNAHGMVMDQPKAALMLRR 477

RESULT 12
CPVL_HUMAN
ID CPVL_HUMAN STANDARD; PRT; 476 AA.
AC Q9H3G5; Q9NB17; Q96AR7; Q9HB41;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable serine carboxypeptidase CPVL precursor (EC 3.4.16.-)
DE (Carboxypeptidase, vitellogenic-like) (Vitellogenic carboxypeptidase-
DE like protein) (VCP-like protein).
GN CPVL OR VLP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21295045; PubMed=11401439;
RA Mahoney J.A., Ntcolosi B., DaSilva R.P., Gordon S., McKnight A.J.;
RT "Cloning and characterization of CPVL, a novel serine
RT carboxypeptidase, from human macrophages.";
RL Genomics 72:243-251(2001).
RN [2]
SEQUENCE FROM N.A.
RA Cho J.-J., Baik H.-H.;
RT "Cloning of VCP-like protein expressed in human heart and placenta.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
PC TISSUE-Placenta;
RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,
RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,
RA Nagahari K., Sugano S., Isogai T.;
RT "HRI human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]

```

SEQUENCE FROM N.A.
 RP TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McWan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grumwood J.J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.J., Skalska U., Smalilus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: May be involved in the digestion of phagocytosed
 CC particles in the lysosome, participation in an inflammatory
 CC protease cascade, and trimming of peptides for antigen
 CC presentation.
 CC -!- TISSUE SPECIFICITY: Expressed in macrophages but not in other
 CC leukocytes. Abundantly expressed in heart and kidney. Also
 CC expressed in spleen, leukocytes, and placenta.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF106704; AAG37991.2; -
 CC EMBL; AF282617; AAG14348.1; -
 CC EMBL; AK075433; BAC11618.1; -
 CC EMBL; BC016838; AAH16838.1; -
 CC HSSP; P10619; IIVY.
 CC MEROPS; S10.003; -.
 CC Genew; HGNC:14399; CPVL.
 CC InterPro; IPR000379; Ser_estr_ssite.
 CC InterPro; IPR001563; Serine_carbpept.
 CC Pfam; PF00450; serine_carbpept; 1.
 CC PRINTS; PR00724; CRBOXPFASEC.
 CC ProDom; PD001189; Serine_carbpept; 1.
 CC PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 CC PROSITE; PS00560; CARBOXYPEPT_SER_HIS; FALSE_NEG.
 CC Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 ? POTENTIAL.
 FT CHAIN ? 476 PROBABLE SERINE CARBOXYPEPTIDASE CPVL.
 FT ACT_SITE 204 204 BY SIMILARITY.
 FT ACT_SITE 388 388 BY SIMILARITY.
 FT ACT_SITE 448 448 BY SIMILARITY.
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 25 25 R -> H (IN REF. 2).
 FT CONFLICT 284 284 L -> F (IN REF. 3 AND 4).
 FT CONFLICT 287 287 F -> L (IN REF. 3).
 FT CONFLICT 398 398 H -> R (IN REF. 3 AND 4).
 FT CONFLICT 422 422 F -> L (IN REF. 2).
 FT CONFLICT 435 435 A -> V (IN REF. 3 AND 4).
 FT CONFLICT 438 438 F -> S (IN REF. 2).
 CC SEQUENCE 476 AA; 54110 MW; 2D966683A4F3FD01 CRC64;

Query Match 12.4%; Score 359; DB 1; Length 476;
 Best Local Similarity 26.7%; Pred. No. 6.5e-18;
 Matches 130; Conservative 62; Mismatches 178; Indels 116; Gaps 21;
 QY 47 PG--VKSYSGY--VDTSPESTHFFWFEARHNPETAPITLWNGGPGSDSLIGLFEELGP 102
 DB 67 PGLNMSYAGFLIVNKYNSLFFWFEPAQIQEDAPVVLWLOGGPGSSMEGLFVEHGP 126
 QY 103 CHVNSTDDYINPHSNWNEVSNLLFLSPLQGVGSYSSTVDGNSINPVYGVVENSFAGVQG 162
 DB 127 YVTSNNTLRDRDPFWTTILSMLYIDNPVGTGFSFTDDTHG----- 167
 QY 163 RYFTIDATLDTNLAARAAWEILQGLSLDLSRVQSKDFSLWTSYGGHYGPAFFN 222
 DB 168 -----YAVNEDDVARLDYSLIQFF-----QTFPEYKNDFFVIGESTAGKVPVIAH 215
 QY 223 HEYEQNERIANGSVNGVQLNFNSIGIINGIIDEAICAPYYPPEFVANNIYGIKAVNETVYN 282
 DB 216 LIHSLN-----PVREVKINLNGTAGDYSDEPSIIGVYAEF-----LYQIGLLDEKQKK 265
 QY 283 YMFANOMPNGCQQLISTCKQTN---RTALADYALCAEATN---MCRDNVEG--PYAFAG 335
 DB 266 Y--FQKQ---CHECHIRKQNWLEAFELDKLLDGLTSDPSYFQNVGTGCSNYNFL- 318
 QY 336 RGVYDIRHPYDDPTPPSYNKNFLAKDSVMDAIGNVINYNTQSNNDVY--YAFQOTGDFVWP 393
 DB 319 -----RCTEPEDQLYYVYKFLSLPEVROAIHVG--NOTFNDGTIVVEKYRLREDIVQSVKP 369
 QY 394 NFIEDLEIILALPVRVSLIYGDADYI-----CNWFGSQAVSLAANYSQAAQ-- 439
 DB 370 WLTEIMNN-----YKVLIIYQCQLDIIVAAALTEHSLMGMDWKGSQ-----EYKKAEEKV 418
 QY 440 ---FRS-----AGYTPKLVNGVEYGETREYGNFSTRVVEAGHEVYVYOPTASLOLFNRTI 492
 DB 419 WKIFKSDSEVAGY-----IRQAGDFHOVIIRGGSHILFYDQPLRAFDMINRFI 466
 QY 493 F--GWD 496
 DB 467 YGKGWD 472
 RESULT 13
 CP22_HORVU
 ID CP22_HORVU STANDARD; PRT; 436 AA.
 AC P55748;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Serine carboxypeptidase II-2 precursor (EC 3.4.16.6) (CP-MII.2)
 DE (Fragment).
 DE CAP;2-2.
 GN Hordeum vulgare (Barley).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
 OC Triticeae; Hordeum.
 OC NCBI_TaxID=4513;
 OX [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=cv. Alexis; TISSUE=Grain;
 RX MEDLINE=94336715; PubMed=7520177;
 RA Dal Degun F., Rocher A., Cameron-Mills V., von Wettstein D.;
 RT "The expression of serine carboxypeptidases during maturation and
 RT germination of the barley grain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:8209-8213(1994).
 CC -!- CATALYTIC ACTIVITY: Preferential release of a C-terminal arginine
 CC or lysine residue.
 CC -!- SUBUNIT: CARBOXYPEPTIDASE II IS A DIMER, WHERE EACH MONOMER IS
 CC COMPOSED OF TWO CHAINS LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE GERMINATING EMBRYO. LOW
 CC LEVELS IN THE DEVELOPING ALEURONE AND EMBRYO. ALSO FOUND IN THE
 CC ROOTS AND SHOOTS OF THE GROWING SEEDLING.
 CC -!- PTM: THE LINKER PEPTIDE IS ENDOPROTEOLYTICALLY EXCISED DURING

CC ENZYME MATURATION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X78878; CAB59202.1; -.
 CC HSP: P08819; IWHI.
 CC InterPro: IPR000379; Ser_estrs_site.
 CC Pfam: PF00450; serine_carbpept; 1.
 CC PRINTS: PR00724; CRBOXYPTASEC.
 CC ProDom: PD001189; Serine_carbpept; 1.
 CC PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
 CC PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
 CC KW Hydroxylase; Carboxypeptidase; Glycoprotein; Zymogen; Multigene family.
 CC NON_TER 1
 CC CHAIN <1 256
 CC PROPEP 257 270
 CC CHAIN 271 436
 CC ACT_SITE 149 149
 CC ACT_SITE 350 350
 CC ACT_SITE 403 403
 CC DISULFID 56 313
 CC DISULFID 217 229
 CC DISULFID 253 281
 CC CARBOHYD 107 107
 CC SEQUENCE 436 AA; 48952 MW; E0F82D97E0C34DC9 CRC64;
 CC
 CC Query Match 12.2%; Score 353.5; DB 1; Length 436;
 CC Best Local Similarity 25.7%; Pred. No. 1.4e-17;
 CC Matches 123; Conservative 76; Mismatches 173; Indels 107; Gaps 23;
 CC
 CC 52 YSGYVDTSP--SHTFWFFEARHNPTAPITLWNGPGSDSL-IGLFEELGPGCHVNST 108
 CC 16 YAGYVTSERDGAALFYWFEEAARFASKPALLWNGPGGSSIAFGVGEVGFPHVAD 75
 CC
 CC 109 FDD-YINPHSWNEVSNILFLSQPLGVGFSDTVGDSINPVTGVVNSSAGVQGRPTI 167
 CC 76 GKGVHNPYSWNOVANILFLDSPGVGYSYNT-----SADILSN-----GDRTA 121
 CC 168 DATLIDTTLAARAAWEILQGLFLSLGFLSDRSVQSKDFSLWTSYSGHYGPAFFNHYEQ 227
 CC 122 KDSLVELTK-----N-LERF-----PQYKEREFLYLTGSYAGHYVPQLAQAIKRH 165
 CC 228 NERIANGSVNGVNLNSGLIINGIIDEAIQAPYPEFAFVANNYTGKAVNETVYNYMKFA 287
 CC 166 HEATGDKSI-----NLKGYMVGNAITDDF--HDHYGIFQYMTTGL--ISDQYKLLNF 216
 CC 288 NQMPNCC--QDLISTCKQNRVALADYALCAEATNCRNVEGP--YYAFAG----- 335
 CC 217 -----CDRESFHTSPQDK--ILDIA--STEAGNIDSYISFTPTCHSFSASSRNKVKR 267
 CC 336 -RGVDIRHYDDPTPPS---YNNKFLAKDS--VMDAIG-----VNINYTSNND 379
 CC 268 LRSVGKMGEGYDPCTEKHSIVYENLHVQKALHVPVIGKSKWETCSEVINTNWKDCRS 327
 CC 380 VYVAFQGTDFVWPNEIDLEELLALPVRSYLIYGADYICNWFPGQAVSLAANYSOAAQ 439
 CC 328 VLIHY-----HELIQGLRIWMPSSGTDV-----IPVTSYSDA 364
 CC 440 FRSGAGYTPLVK-----NGVEYGETREYCNFSFTRVYEGAGHEVPYQPIASLQFLNRTIFG 494
 CC 365 LXLPTTPWHAWYDDGEGVGGWYQGYKGLNFVIRVGAGHEVPLHRPKQALTLIKSLFIAG 423

RESULT 14
 YSS2_CAEEL

ID YSS2_CAEEL STANDARD; PRT; 470 AA.
 AC Q09991;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative serine carboxypeptidase K10B2.2 precursor (EC 3.4.16.-).
 GN K10B2.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
 OC Rhabditiidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Miller N.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U28730; AAA68259.1; -.
 CC PIR: T16606; T16606.
 CC HSP: P10619; IIVY.
 CC MEROPS: S10.002; -.
 CC WormPep: K10B2.2; CE02009.
 CC InterPro: IPR000379; Ser_estrs_site.
 CC InterPro: IPR001563; Serine_carbpept.
 CC Pfam: PF00450; serine_carbpept; 1.
 CC PRINTS: PR00724; CRBOXYPTASEC.
 CC ProDom: PD001189; Serine_carbpept; 2.
 CC PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
 CC PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
 CC KW Hypothetical protein; Hydrolase; Carboxypeptidase; Glycoprotein;
 CC Signal.
 CC SIGNAL 1 19
 CC CHAIN 20 470
 CC ACT_SITE 169 169
 CC BY SIMILARITY.
 CC ACT_SITE 380 380
 CC BY SIMILARITY.
 CC ACT_SITE 441 441
 CC BY SIMILARITY.
 CC CARBOHYD 132 132
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 316 316
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 396 396
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 470 AA; 53158 MW; CCC2DACB75EF30FC CRC64;
 CC
 CC Query Match 12.2%; Score 352.5; DB 1; Length 470;
 CC Best Local Similarity 26.4%; Pred. No. 1.8e-17;
 CC Matches 131; Conservative 60; Mismatches 179; Indels 127; Gaps 20;

QY 46 TPGVKSYGVVDTPSPESHTEFFWFFEARHNPTAPITLWNGPGSDSLIGLFEELGPGCHV 105
 Db 35 TPDFHYSGYIRAWTDKYLHWLTESRSRAPTDQPLVLWNGPGGSSLDGLIELGPGFHV 94
 QY 106 NSTFDD-YINPHSWNEVSNILFLSQPLGVGFSDTVGDSINPVTGVVNSSAGVQGRY 164
 Db 95 KDFGNSIYYNEYAWNKFANVLFESAGVGYSTINFLTVSDDEVSLN----- 144
 QY 165 PTIDATLIDPTNLAAEAAWEILQGLFLSLGFLSDRSVQSKDFSLWTSYSGHYGPAFFNHP 224
 Db 145 -----YMALDLFLSKPEYKGR---DFWITGSYAGVIPTL----- 178
 QY 225 YEONERIANGSVNGVNLNSGLIINGIIDEAIQAPYPEFAFVANNYTGKAVNETV 280
 Db 179 ---AVRILNDKKFP--NEKGAIGNGALNFPNNYTMVFPY-----YYHALVRDDL 225
 QY 281 YNYMKEANQMPNGCQDLISTCKQNRVALADYALCAEATNCRNVEGPPYAFAGRGVYD 340
 Db 226 YN-----DIARNCCNNIGTCD-----IYKFFDPN--CRDKV---INALDGTNELN 267

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2003, 22:37:42 ; Search time 78 Seconds
(without alignments)
1776.593 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555
Perfect score: 2887
Sequence: 1 LPOSTPASVGRRLPKNPFG.....HTQSSVPLPFAISMSSVGMA 537

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2884	99.9	555	3 Q96VZ9	Q96vz9 aspergillus
2	746.5	25.9	460	3 Q12569	Q12569 absidia zyc
3	601	20.8	552	3 Q96VC4	Q96vc4 emericella
4	596.5	20.7	541	3 Q94152	Q94152 pichia anqu
5	595.5	20.6	537	3 Q14414	Q14414 pichia anqu
6	479	16.6	507	10 Q8VWQ0	Q8vwq0 gossypium h
7	473	16.4	429	10 Q8GV71	Q8gv71 oryza sativ
8	466.5	16.2	505	10 Q9FFB0	Q9ffb0 arabidopsis
9	461	16.0	501	10 Q9XH61	Q9xh61 matricaria
10	460	15.9	508	10 Q8L6A7	Q8l6a7 theobroma c
11	439	15.2	510	10 Q9LXH4	Q9lxh4 arabidopsis
12	438	15.2	510	10 Q93ZC3	Q93zc3 arabidopsis
13	413.5	14.3	2105	5 Q17679	Q17679 caenorhabdi
14	406.5	14.1	499	10 Q9FMX9	Q9fmx9 arabidopsis
15	385	13.3	490	10 Q9FYF7	Q9fyp7 oryza sativ
16	383	13.3	482	10 Q9SV04	Q9sv04 arabidopsis

17	376.5	13.0	524	10 Q8GTK2	Q8gtk2 oryza sativ
18	363	12.6	480	10 Q9LEY1	Q9ley1 arabidopsis
19	360	12.5	437	10 Q9FWG1	Q9fwg1 oryza sativ
20	360	12.5	482	10 Q9FRJ0	Q9frj0 oryza sativ
21	359	12.4	487	10 Q9SV03	Q9sv03 arabidopsis
22	357.5	12.4	479	10 Q949Q7	Q949q7 arabidopsis
23	355.5	12.3	360	10 Q9M450	Q9m450 cicer ariet
24	353	12.2	470	10 Q22732	Q22732 arabidopsis
25	349	12.1	474	11 Q9D2D1	Q9d2d1 mus musculu
26	348	12.1	452	10 Q9ZQO0	Q9zqo0 arabidopsis
27	344	11.9	465	10 Q04084	Q04084 arabidopsis
28	344	11.9	494	10 Q9FEU4	Q9feu4 pisum sativ
29	344	11.9	501	10 Q9SV02	Q9sv02 arabidopsis
30	342.5	11.9	471	10 Q8L9Y0	Q8l9y0 arabidopsis
31	342.5	11.9	473	10 Q9SGA9	Q9sga9 arabidopsis
32	342.5	11.9	669	10 Q8L9Y6	Q8l9y6 oryza sativ
33	340	11.8	512	5 Q76725	Q76725 caenorhabdi
34	337.5	11.7	425	10 Q65568	Q65568 arabidopsis
35	333.5	11.6	510	3 Q60123	Q60123 schizosacch
36	333	11.5	487	10 Q9Z0G3	Q9zu03 arabidopsis
37	332	11.5	465	10 Q9M099	Q9m099 arabidopsis
38	331	11.5	497	10 Q8L7H2	Q8l7h2 arabidopsis
39	330.5	11.4	472	10 Q9LSM9	Q9lsm9 arabidopsis
40	330.5	11.4	474	10 Q82228	Q82228 arabidopsis
41	330	11.4	445	5 Q814E3	Q814e3 caenorhabdi
42	330	11.4	1203	5 Q45916	Q45916 caenorhabdi
43	328.5	11.4	548	4 Q9BR08	Q9br08 homo sapien
44	328	11.4	504	10 Q9LSV8	Q9lsv8 arabidopsis
45	327.5	11.3	2338	5 Q94269	Q94269 caenorhabdi

ALIGNMENTS

RESULT 1

Q96VZ9 PRELIMINARY: PRT: 555 AA.
AC Q96VZ9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
ET 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Carboxypeptidase S1 precursor (EC 3.4.16.6).
GN CPI.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurothiales; trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK3;
RA van den Broek P.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF394242; AAK77166.1; -
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
KW Carboxypeptidase; Hydrolase; Signal.
FT SIGNAL 1
FT POTENTIAL 18
SQ SEQUENCE 555 AA; 61168 MW; 456863B0CEB55222 CRC64;

Query Match 99.9%; Score 2884; DB 3; Length 555;

Best Local Similarity 99.8%; Pred. No. 1.1e-195;

Matches 536; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPGSTPASVGRQLPKNPFGVTKLTANNVILRYKEPGAEGVCETPGVKSYSGYVDSP 60

|||||

Db 19 LPGSTPASVGRQLPKNPFGVTKLTANNVILRYKEPGAEGVCETPGVKSYSGYVDSP 78

|||||

Qy 61 ESHTFFWFEEARHNDETAPITLWNLGGPGSDSLIGLFEELGPHCHVNSFDDYINPHSWNE 120

|||||

Db 79 ESHTFFWFEEARHNDETAPITLWNLGGPGSDSLIGLFEELGPHCHVNSFDDYINPHSWNE 138


```

Db 166 K-----NSEGHINLKGAIGNGLTDPATQYKAYTDYSLD-----NGLTKSOFNFI---N 213
QY 289 QMPNCGQDLISTCKQTNR-TALADYALCAEATNMCRDNVEGPPYAFAGGVYDIRHP--- 344
Db 214 KIVPTCELAIAKLCGTSGTISGLGAYVVC-----NLIFFSIE---TIGKKNYYDIRKPCVG 266
QY 345 ---YDDPTPPSYNKFELAKDSVMDAIGV-NINYTQSNNDVYAFQOTGDFVWPNTIE-DL 399
Db 267 SLCYD-----LSNMEKFLQKSVRESLGVGDIOFVSCSTVTOAML-----LDMNRLEVGI 318
QY 400 BEILALPVRSYLIYGDADYICNWFEGGQAVSLAANYSOAQRSGYTPKLVNGVYGETR 459
Db 319 PELLENDIKVLIYAGEYDLICNWLGNRWNSWMSWKEAFVSSREBFTVDGKAGILK 378
QY 460 EYGNFSTRVYAEAGHEVPYIOPIASLOLFNRTIFG 494
Db 379 SYGLPSFLKHVDAGHMVPMQPKVALEMLMRWTSG 413

RESULT 8
Q9FFB0
ID Q9FFB0 PRELIMINARY; PRT; 505 AA.
AC Q9FFB0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Serine carboxypeptidase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=97411969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pl clones.";
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005243; BAB10619.1;
DR HSSP; P10619; IIVY.
DR MEROPS; S10.009; -.
DR InterPro; IPR001563; Serine_carpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
KW Carboxypeptidase.
SQ SEQUENCE 505 AA; 56543 MW; 6E74CD351090B099 CRC64;

Query Match
Best Local Similarity 16.2%; Score 466.5; DB 10; Length 505;
Matches 132; Conservative 75; Mismatches 178; Indels 81; Gaps 15;
QY 47 PGVKS---YSGY--VDTSPESHITFFWFFEARHNPETAPITLWNGPGSDSLIGLFEELG 101
Db 91 PSVOEFCHYAGYSILPHSKAKMFYFFESR-NKTTDPVVYIWLIGGPGCSSVAMFYENG 149
QY 102 PCHVNSTFDYINPHSWNEVSNLLFLSQPLGVGFSDYTDVGDGINPVTGVWNSFAGVQ 161
Db 150 PFKISKDLSLYWDFDQWVKYSNIYYDQPVGTGFSYTSDESIRNDEGV----- 199
QY 162 GRYPTIDATLIDTNLAEEAAWEILQGLSLSDSRVQSKDFSLWTSYGGHYGAPFF 221
Db 200 -----SNDLYDLQAFKHEP-----KVKNDFFITGESYAGHYIPALA 238
QY 222 NHYEQNERIANGSVNGVQLNENSLGLINGIIIDEAIOAPYYPEFVNNYVIGIKAVNEIVY 281

```

```

Db 239 SRVHSGNKK-----KEGIPINLKGAIGNGLNPEIQYGAIDYALQ-----MKLISESDH 289
QY 282 NTMKEANOMPNGCQDLISTCKQTNRTALADYALCAEATNMCRDNVEGPPYAFAGGVYDI 341
Db 290 ESLK-----QDYVEQNTIKKCSGLGGVLCDSAVEVCSIFENKIVAKKSGUNYDI 340
QY 342 RHP-----YDDPTPPSYNKFELAKDSVMDAIGV-NINYTQSNNDVYAFQOTGDFVWP 394
Db 341 RKKCVGSLCYDP---SRMEIFLNKENVRKALGVGDIFVSCSSIVY-----DAMIED 389
QY 395 FIEDLEEILALPVRSY-----LIY-GDADYICNWFEGGQAVSLAANYSOAQRSGYTP 448
Db 390 WMONLE--VKPISLVNDGILNLYAGYEDLICNWLGNRWVDMWNSGOKGFGSAKNVSF 447
QY 449 KVNVEGETREYGNFSTRVYAEAGHEVPYIOPIASLOLFNRTIFG 494
Db 448 LVDGKEAGLLKNHGLTFLKYNAGHMVPMQPKASLOLQNMWMOG 493

RESULT 9
Q9XH61
ID Q9XH61 PRELIMINARY; PRT; 501 AA.
AC Q9XH61;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Serine carboxypeptidase.
OS Matricaria chamomilla.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Anthemideae; Matricaria.
OX NCBI_TaxID=98504;
RN [1]
RP SEQUENCE FROM N.A.
RA Kohchi C., Yasuda H., Hirata T.;
RT "Isolation of a cDNA encoding for a carboxypeptidase, having leucine
RT zipper structure at the N-terminal region, from the cultured shoot
RT primordia of Matricaria chamomilla.";
RL Plant Biotechnol. 16:409-412(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Kohchi C., Yasuda H., Hirata T.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF141384; AAD42963.2; -.
DR HSSP; P00729; IYSC.
DR MEROPS; S10.009; -.
DR InterPro; IPR001563; Serine_carpept.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00450; serine_carpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine_carpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
KW Carboxypeptidase.
SQ SEQUENCE 501 AA; 55973 MW; DFD8F8DB41880A66 CRC64;

Query Match
Best Local Similarity 16.0%; Score 461; DB 10; Length 501;
Matches 131; Conservative 76; Mismatches 188; Indels 72; Gaps 15;
QY 39 AEGVCETTPGVKYSGY--VDTSPESHITFFWFFEARHNPETAPITLWNGPGSDSLIGL 96
Db 80 SDGV--TVEDLGHGAGYQIQHSHAAKMFYFFESRNNKKD-PVWILTGPGCSSLEAL 136
QY 97 FEELGCHVNSTFDDYINPHSWNEVSNLLFLSQPLGVGFSDYTDVGDGINPVTGVWNS 156
Db 137 FYENGPFKIADNMTLVWNEYGQDQANLIYVDQGTGTGFSYSSDKDIRHDEQGV----- 191
QY 157 FAGVQGRYPTIDATLIDTNLAEEAAWEILQGLSLSDSRVQSKDFSLWTSYGGHY 216
Db 192 -----SDDLYDLQAFKHEP-----VDNDFYITGESYAGHY 225

```

```
QY 217 GPAFFNHFYEQRNRIANGSVNGVQLNFNSLGIINGIIDEAIAQAPYYPFAVNNYTGKAV 276
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 226 IFAIAARVHGNGK-----AKEGHIHLNKGFICNGLTDPALQYQAYDYALD---MGLI 276
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 277 NETVYNYMKFANPMPGCGDLISTCKOTNR--TALADYALCAEATNMCNRDVEGPPYAFAG 335
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 277 KESQY---KRINLIVLPCAIAIKLGGTGTGTCISCMASYFVC-----NAIFTGIMALAG 335
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 336 R-GVYDIRHP-----YDDETPPSYNNKFLAKDSVMDAIGV--NINYTQSNNDVYXAFQOT 387
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 326 NINYDIRKECVTSMCYDF-----SDMETLNLKKSVRQALGVGDIEFVSCSTTIVTAMLVD 381
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 GDFWPNFTEDLEILLALPVRSLIYGDADYICNWFQGAQVSLAANTSOAAQFSRAGYTP 447
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 ---LNRNLKAGIPLELLEDGKMLVYAGEYDVICNLGNSWVHAMENWGKEQNALSEAP 438
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 448 LKYNVEYGTREYGNFSFTRVYAGHEVYYPQIASLQFNRTIFGW--DIAFGQKKI 504
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 439 FEVDGSEAGLLKSGPLSELKLVHDAGHMVPMQPKAALEMLKRWMDG 485

RESULT 10
Q8L6A7
ID Q8L6A7 PRELIMINARY; PRT; 508 AA.
AC Q8L6A7
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase type III.
GN CP-III.
OS Theobroma cacao (Cacao) (Cocoa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.
OX NCBI_TaxID=3641;
RN [1]
RP SEQUENCE FROM N.A.
RA Tazi H., McCarthy J., Bucheli P., Laloi M.;
RT "Molecular characterisation of a type-III carboxypeptidase (cacao Cp-
III) from Theobroma cacao seeds.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ313514; CAC86383.1; -
DR InterPro; IPR001563; Serine_carbpept.
DR pfam; PF00450; Serine_carbpept.
DR PRINTS; PR00724; CROXYPTASEC.
DR ProDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE 508 AA; 56521 MW; 68B63707DC686BFE CRC64;

Query Match 15.9%; Score 460; DB 10; Length 508;
Best Local Similarity 27.2%; Pred. No. 2.9e-24;
Matches 146; Conservative 88; Mismatches 206; Indels 96; Gaps 21;

QY 1 LPGSTPASVGRROQ-----PKNPTGVKTLTANNVTI-----RYKEPGAEGVCE 44
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 34 LGGSFSPSIHAKLLRELNLFPKEVNV---VDGQVSLPDSRLVREKRFKNL----- 85
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 45 TTPGVKVS-----YSGY---VDTSPSHFTFFWFHFAHNPFPAPITLWNLGPGSDSLGL 96
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 86 AVPGVGSVEDLGHAGYKLANSHDARMFYFFFSR-NSKKDPVVIWITGGPGCSSEAL 144
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 FEELGPCVNSTFDDYINPHSNWEVSNLELISQPLGVGFSYSDTVGDSINPTGVVENS 156
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 145 FYENGPFPTIAENMSLIWNOYGDMAENLLYVQPIGTCFSYS-----SD 188
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 FAGVGQRYPTIDATLIDTNTLAAEAWEILQGLSLPDSRVQSKDFSLWTEYSGHY 216
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 RRDIRHNEDEVSNDLYD-----FLQAFPAEHPEF-----EKNDFYITGESYAGHY 233
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 217 GPAFFNHFYEQRNRIANGSVNGVQLNFNSLGIINGIIDEAIAQAPYYPFAVNNYTGKAV 276
```

```
Db 234 IPFAARVHGNGK-----AKDGHILNKGFALGNGLTDPALQYKAYDIYALD---MGVI 284
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 277 NETVYNYMKFANPMPGCGDLISTCKOTNR--TALADYALCAEATNMCNRDVEGPPYAFAG 335
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 285 KKSQYNR---NKLVPVCENAIKLGTDGTGTCISCMASYFVC-----NAIFTGIMALAG 333
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 336 -RGVYDIRHPDPTPPSYNN--KFLAKDSVMDAIGV--NINYTQSNNDVYXAFQOTGDFV 391
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 334 DTNVDIRTKCEGSLCYDFSNMETFNLQESVRDALGVGSDIFVSCSTVYQOAMLVD--- 389
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 WPNFIE--DLBEILLALPVRSLIYGDADYICNWFQGAQVSLAANTSOAAQFSRAGYTKV 450
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 390 WMRNLEVGIPALLEDGKMLVYAGEYDVICNLGNSWVHAMENWGKEQFVASEVFPV 449
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 451 NGVYGTREYGNFSFTRVYAGHEVYYPQIASLQFNRTIFGW--DIAFGQKKI 504
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 DGSAGVLRVTHGPLGFLKLVHDAGHMVPMQPKAALEMLKRWTKGTLSEADSEKLV 505

RESULT 11
Q9LXH4
ID Q9LXH4 PRELIMINARY; PRT; 510 AA.
AC Q9LXH4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Carboxypeptidase precursor-like protein (AT3g45010/F14D17_80).
GN F14D17_80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unseld M., Mewes H.W.,
RA Rued S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsids sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Koesema E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA Arabidopsids cDNA clones."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RL "Arabidopsids ORF clones."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL353992; CAB89316.1; -
DR EMBL; AY050427; AAK91443.1; -
DR EMBL; AY093993; AAM16254.1; -
DR HSSP; P00729; 1YSC.
DR MEROPS; S10.009; -.
DR InterPro; IPR001563; Serine_carbpept.
```

DR	InterPro: IPRO00379; Ser_estr_site.	
DR	Pfam: PF00450; serine_carbpept; 1.	
DR	PRINTS: PR00724; CRBOXPIASEC.	
DR	ProDom: PD001189; Serine_carbpept; 1.	
DR	PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.	
DR	PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.	
DR	SEQUENCE 510 AA; 5696 MW; 134BC30FAE64BF0A CRC64;	
Query Match	15.2%; Score 439; DB 10; Length 510;	
Best Local Similarity	26.6%; Pred. No. 8.9e-23;	
Matches 123; Conservative 81; Mismatches 189; Indels 69; Gaps 14;		
QY	46 TPGVKSY---SGY--VDTSPESHWFPEFEARHNPETAPITWLKNGPGCDSLGLPEEL 100	
DB	93 SPVQDGHGHHGYKLPNSKAARPIYFFFEKSTN-KADPVYIWLITGPGCSSEIALEYEN 151	
QY	101 GPCHVNSTFDDYINPHSWNEVSNLLFSLQPLGVGFSYSDTVGDSINPVTGVVENSFAGV 160	
DB	152 GPTVSNNSLSWNEFGWDKASNLIIYDOPVGTGFSYTSQSDLRHDEDGV----- 202	
QY	161 QGRYPTDAVLITDTNLAAEAAEWELLQGFISGLPSLDSRVOSKDFSLWTESYGGHYGPAF 220	
DB	203 -----SNDLYDFLOAFKEHP-----QFYKNDFTIGESYAGHYTPAL 240	
QY	221 FNHFEQNERIANGSVGNLNSLGTINGILIEATQAPYYPEFVANNVTGIKAVNEIV 280	
DB	241 ASRVHRGNK-----NKGETHNLKGFPAIGNLTNPETQYQAYAYALDMNLITQSDHNL 295	
QY	281 YNYMKFANOMPGQDLISCTQTNRTALADYA---LCAEATNMCRONVEGPPYAFAGF- 336	
DB	296 NRY-----YATQQSICEKSDAGGEDACASSYTV-CNNIFQKIMDIAGNV 340	
QY	337 GVYDIRHPYDDPTPPSYN--KFLAKUSVMDAIGV-NINTYOSNDVYIYAFQQTGDFWVP 393	
DB	341 NYIDVRKQCEGSLCYDFSNMENFLNQKSVRKALGVGDIIEFVSCSTAYVEAMQMD----WM 396	
QY	394 NFIE-DLEEIIIALPVRVSLIYGDADYICNFFGGQAVSLAANYSOAAQFRSAGIYPLKVG 452	
DB	397 RNLVEGPALLQDGIKLLVYAGEVDLLCNLWLGSKWHEMSQCKEFVAAATVFFHVDN 456	
QY	453 VEYGETREYGNFSPTRVYEGAGHEVYVQPIASIQLFNRTIFG 494	
DB	457 KEAGLMKNYSGLTFLKVDHAGHVPMPQPPRAAQLOMLNWMOG 498	

```

RESULT 12
Q93ZC3
ID Q93ZC3 PRELIMINARY; PRT; 510 AA.
AC Q93ZC3:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE A3q945010/F14D17_80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carinci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RA "Arabidopsis cdna clones.";
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RL
DR EMBL: AY057639; AAL15270.1; -
DR InterPro: IPR001563; Serine_carbpept.
DR InterPro: IPR000379; Ser_estrs_site.

```

```

DR Pfam: PF00450; serine_carbpept; 1.
DR PRINIS; PD00724; CRBOXYPTASEC.
DR proDom; PD001189; Serine_carbpept; 1.
DR PROSITE; PS00160; CARBOXYPEPT_SER_HIS; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
SQ SEQUENCE      510 AA; 56837 MW; 704BC00CAE564BF08 CRC64;

Query Match          15.2%; Score 438; DB 10; Length 510;
Best Local Similarity 26.6%; Pred. No. le-22;
Matches 123; Conservative 80; Mismatches 190; Indels 69; Gaps 14

QY      46 TPGVKSY---SGY---VDTSPESHTEFWFEARHNPEIAPITLWLNGCGSPDSLGLGFEEL 100
       || : | : : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      93 SPVSQDFGHAGYYKLPSNKAARMYFFESRTN-KADPVVILWLTGGCCSEALFYEN 151

QY     101 GPCHVNSTFDDYINPHSNNEYSNLLFLSOPLGVGFSYSDTVDGSTNPVTGYVNSSFGAV 160
       || | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     152 GFPVTSSNNSSLSWNIEFGWKASNLIVDPQVGTGFESTSDQSDLRHDEGV----- 202

QY     161 QGRYPTIDATLITDTLNAAEAWEILOGLFLSLPDSLRVQSKDSLESYGVGHYPAF 220
       || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     203 -----SNDLYDFLQAFFKEHP-----QFVKNDYIYGESYAGHYIAL 240

QY     221 ENHFYEQNERTANGSVGVQLNFNSGLINGLIIDEAIOAPYPERFAVNTYGIKAVNETV 280
       || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     241 ASRVHRGNK-----NKGETHNLKGFALGNGLTNPETIQGYAYDALDMNLITQSDHNL 295

QY     281 YNMKFANOMPNGCODLSTCKQTNRKTALADYA---LCAEATNMCNRDNVEGPYYAFAGR- 336
       || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     296 NRY-----YATCOOSTIKCSADGEDCACASYTVC--NNIFOKIMDIAGNV 340

QY     337 GVYDIRHPYDDETPPSYNN--KFLAKOSVMDAIGV-NINYTQSNNDNYVAFQOTGDWFVP 393
       || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     341 NYDVVRKOCEGSCLCYDFSMMENFLNQKSVRKALGVGEIEFYVSCSTAIFYEAMQMD----WM 396

QY     394 NFIE-DLEEEITALPRVRSIIYGDADYICNWTFGGQAVSLAANYSOAAQFRSAGYTPLKYN 452
       || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     397 RNLEVGIPALLQDGKILLVYAGEYDLCNWLGNKKWHHEMSQKEFVAATVPFFHVGN 456

QY     453 VEYGETREYGNFTSTRVYEACHEGVPPYQPIASLQLFNFTIFG 494
       || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     457 KEAGLMKNYSLEIFLKVHDAGHYVPMQGPKAALQMQLNWGMQ 498

```

```

RESULT 13
Q17679 PRELIMINARY; PRT; 2105 AA.
AC Q17679;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Y1684A.2 protein.
GN Y1684A.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Thomas K.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z70203; CAA94110.1; -.
DR EMBL; AL023825; CAA94110.1; JOINED.
DR EMBL; Z93339; CAA94110.1; JOINED.
DR EMBL; Z93339; CAB07544.1; -.
DR EMBL; AL023825; CAB07544.1; JOINED.
DR EMBL; Z70203; CAB07544.1; JOINED.
DR EMBL; AL023825; CAA19443.1; -.
DR EMBL; Z93339; CAA19443.1; JOINED.
DR EMBL; Z70203; CAA19443.1; JOINED.
DR HSP; P10619; ITV.
DR Wormprep; Y1684A.2; CE21374.
DR InterPro; IPR001563; Serine_carbpept.

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: September 16, 2003, 22:52:38 ; Search time 6324 Seconds
(without alignments)
3473.825 Million cell updates/sec
Title: US-09-712-338-2_COPY_19_555
Perfect score: 2887
Sequence: 1 LFGSTPASVGRRLPKNPQTG.....HTGSSVPLPTATSMSSVGMGA 537

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO.spool.p/US09712338/runat_16092003_144914_19130/app_query.fasta_1.711
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGL_1.1.4958 @runat_16092003_144914_19130 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pla.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2883	99.9	1662	6	AR129928 Sequence
2	2526	87.5	2245	6	AF394242 Aspergill
3	2338	81.0	1656	6	AX534871 Sequence
4	1983	68.7	3150	6	AX534814 Sequence
5	980.5	34.0	1872	6	AX534872 Sequence
6	977	33.8	3221	6	AX534815 Sequence
7	941	32.6	1665	6	AX534866 Sequence
8	916.5	31.7	1581	6	AX534824 Sequence
9	894	31.0	2940	6	AX534767 Sequence
10	887.5	30.7	3080	6	AX534809 Sequence
11	652.5	22.6	4308	8	D86560 Schizosacch
12	652.5	22.6	37000	8	SPAC19G12
13	636	22.0	1527	6	AX594884 Sequence
14	636	22.0	2016	6	SCYBR139W
15	636	22.0	2027	6	AX536454 Sequence
16	636	22.0	29686	8	SC29711
17	636	22.0	50277	2	AC138524_5
18	630	21.8	1611	6	AX534826 Sequence
19	623	21.6	2002	6	I33983 Sequence 3
20	623	21.6	2002	6	I74375 Sequence 3
21	623	21.6	2002	6	I77239 Sequence 3
22	596.5	20.7	2509	8	AF085063 Pichia an
23	596	20.6	1985	8	YSCARPEPY
24	595.5	20.6	2214	8	PAU67174
25	594.5	20.6	3850	6	E12103
26	594.5	20.6	3850	8	PPPRCLGEN
27	594	20.6	2068	6	I33982
28	594	20.6	2068	6	I74374
29	594	20.6	2068	6	I77238
30	580	20.1	2660	6	AX534769 Sequence 1
31	568.5	19.7	2503	6	A75535 Sequence 2
32	568.5	19.7	2503	6	I70282 Sequence 3
33	548.5	19.0	2280	6	AB051820 Aspergill
34	541.5	18.8	1653	6	AX536918 Sequence
35	527.5	18.3	2632	8	YSCPRCCPY
36	527.5	18.3	15581	8	SC8175
37	523.5	18.1	2632	6	AR129959 Sequence
38	479	16.6	1787	8	AY072822 Gossypium
39	476.5	16.5	1772	8	RICCBP31
40	474	16.4	1551	6	AX12745 Sequence
41	474	16.4	1551	6	AX505988 Sequence
42	474	16.4	1551	8	AY149954 Arabidops
43	474	16.4	1771	8	AY091767 Arabidops
44	470.5	16.3	3298	8	ABGSCFZ
45	461	16.0	1816	8	AF141384 Matricari

ALIGNMENTS

RESULT 1

ARL129928
 LOCUS ARL129928 1662 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 1 from patent US 6187578.
 ACCESSION ARL129928
 VERSION ARL129928.1 GI:14117825
 KEYWORDS
 SOURCE Unknown.
 ORGANISM
 Unclassified.
 REFERENCE 1 (bases 1 to 1662)
 AUTHORS Blinkovsky,A., Berka,R., Rey,M., Golightly,E., Klotz,A., Mathisen,T.,Erik., Dambmann,C. and Brown,K.M.
 TITLE Carboxypeptidases and nucleic acids encoding the same
 JOURNAL Patent: US 6187578-A 1 13-FEB-2001;
 FEATURES
 Location/Qualifiers
 1..1662
 /organism="unknown"
 BASE COUNT 396 a 468 c 398 g 400 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7,36e-218 Length: 1662
 Score: 2883.00 Matches: 536
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.86% Indels: 0
 DB: 6 Gaps: 0
 US-09-712-338-2_COPY_19_555 (1-537) x ARL129928 (1-1662)
 QY 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 20
 DB 55 CTTCAGGAAGTACACCGGCTCGTGGTAGAAGACAGTACCAGAAACCCACCGG 114
 QY 21 ValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu 40
 DB 115 GTCAAGACTCTTACACCGCAACATGTCCATCCGTTACAGGAACCCGGGACAG 174
 QY 41 GlyValCysGlnThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60
 DB 175 GGGCTCTGGAGACTACCCGGGTGCAATCTTACCTCTGGATGATGTCGACACCTCC 234
 QY 61 GluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIle 80
 DB 235 GAGTCCCATACCTCTTCTGTTCTTGAAGCCAGACATAACCCAGAACTGCACCTATC 294
 QY 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeu 100
 DB 295 ACATGTGGTGAATGGTGGCCCTGGAAGCGATCTTTGATCGGTCTCTTCGAAGAGTTG 354
 QY 101 GlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGlu 120
 DB 355 GGCCCTTGCCATGTCATTCGACTTTGATGACTACATCAACCTCCTACTCGTGAACGAG 414
 QY 121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 140
 DB 415 GTCTCCCAATTACTATCTCTGCCAGCCATGGAGTGGGAGTGGCTTTTCATATAGTACG 474
 QY 141 ValAspGlySerIleAsnProValThrClyValValGluAsnSerSerPheAlaGlyVal 160
 DB 475 GTTATGGGTCCATTAACCCGTAACTGGGGTCTGTCGAAATTCAGCTTTTCAGGAGTT 534
 QY 161 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGlu 180
 DB 535 CAGGCGCGGTACCAACCAATGATGCCACTCTGATCGATACTACCAATCTTGGCCGACAG 594
 QY 181 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 200
 DB 595 GCCGCTTGGGAGATCTCTCAAGAGATTCCTTAGTGGACTACCTAGCTTGGACTAGGGTG 654
 QY 201 GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhe 220
 DB 655 CAGTCTAAGGACTTCAGTCTATGGAGCGGAGAGCTATGGAGGGCAGCTATGTCCTGCTTC 714

QY 221 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 240
 DB 715 TTCAATCAUTTTTACGAGCAGAATGAGAGAATGCCAACGTAAGTGTATGTTGTTGTCAG 774
 QY 241 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaPro 260
 DB 775 CTTAATTTCAACTCTCTGGGAATTAATACGGCATCATCGACGAGGGAATCCAGGCCCT 834
 QY 261 TyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal 280
 DB 835 TACTACCTGAATTCGCTGTGAACAATACCTACGGTATCAAGCTGTCAACGAGACGTC 894
 QY 281 TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 300
 DB 895 TACAACCTACATGAAGTTTGCCAAACCAATGCTTCCAGAGATTTGATTTCACCC 954
 QY 301 CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsp 320
 DB 955 TGCAAACAGACAAACCCGCGCATTAAGCTGACTACGCCCTCTCGCCGGAAGCACCAAC 1014
 QY 321 MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 340
 DB 1015 ATGTGCAGGACAAATGTTGAGGGCCATACCTACGCCCTTGTGCTGCTGGTGTGTATGAT 1074
 QY 341 IleArgHisProTyrAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 360
 DB 1075 ATTGGCATCCATGATGATGACCCGACCTCCGCAAGTTATTACAAACAATTTCTGGCAAG 1134
 QY 361 AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 380
 DB 1135 GACTCTGTCTATGGAGCTATCGGGCTCAACAATCAACTACACCCCTCCAAATATGACGTC 1194
 QY 381 TyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 400
 DB 1195 TACTACGCTTTCAGAGAAACAGGCGACTTGTCTGGCCCACTTCACTGAAGACCTCGAG 1254
 QY 401 GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 420
 DB 1255 GAGATCTTGTCTCCCGGTGCTGCTCCCTCATCTATGCGCAGCGCGATTATCATCTGC 1314
 QY 421 AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPhe 440
 DB 1315 AACTGTGTGCGGGTCAAGCGGTTTCCCTCCTCGCACTACTCCCAAGCCGCCAGTTC 1374
 QY 441 ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 460
 DB 1375 CGAAGCGAGGTACACGCCCTGAAAGTCAACGCGCTCGAGTATGGGAAACTCGCGAG 1434
 QY 461 TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTrpGln 480
 DB 1435 TATGTAATTTCTCTTCACTCGGTCTATGAGCAGCGCCAGGAAGTCCCACTACTACCAG 1494
 QY 481 ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 500
 DB 1495 CCCATCGCCTCCCTGCAATTTGTTAACCGGACTATCTTCGGTGGGATATCGCAGAGGC 1554
 QY 501 GlnLysLysIleTrpProSerTyrTrpThrAsnGlyThrAlaThrAlaThrHisThrGln 520
 DB 1555 CAGAAGAAGATCTGGCCCGAGCTACAAGACGAATGGAACGGCTACAGCTACGCATACACAG 1614
 QY 521 SerSerValProLeuProThrAlaThrSerMetSerSerValGlyMet 536
 DB 1615 TCGTCCGTGGCGCTGCGCTACGGCTACCGCATGTCGAGTGTGGTATG 1662
 RESULT 2
 AF394242 2245 bp DNA linear PLN 24-JUL-2001
 LOCUS Aspergillus oryzae strain TK3 carboxypeptidase S1 (cpt) gene,
 complete cds.
 DEFINITION AF394242
 ACCESSION AF394242
 VERSION AF394242.1 GI:15004615
 KEYWORDS

SOURCE Aspergillus oryzae
 ORGANISM Aspergillus oryzae
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 REFERENCE 1 (bases 1 to 2245)
 AUTHORS van den Broek, P.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUN-2001) Bioscience, Nestec S.A., P.O. Box 44,
 Lausanne CH-1000, Switzerland
 FEATURES
 Location/Qualifiers
 source 1..2245
 /organism="Aspergillus oryzae"
 /mol_type="genomic DNA"
 /strain="Tk3"
 /db_xref="taxon:5062"
 <1..>2245
 /gene="cpi"
 join(<1..349,409..513,576..692,743..866,941..956,
 1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)
 /gene="cpi"
 /product="serine carboxypeptidase 1"
 join(1..349,409..513,576..692,743..866,941..956,
 1008..1124,1187..1237,1296..1449,1522..1580,1670..>2245)
 /gene="cpi"
 /EC_number="3.4.16.6"
 /codon_start=1
 /product="carboxypeptidase S1"
 /protein_id="AAK71166.1"
 /db_xref="GI:15004616"
 /translation="MRGEFLSVLPVAAASWALPGSTFASVGRRLKPNPTIGVKTLIT
 ANWTVRKEPGAEGVCTTPGVKSYGYVDTPESHTFFWFPEARHNPTAPTLM
 NGGSDSLGLEELGPCVHNSTFDYINPHSWNEVSNLLELSQPLGVGSYNDYD
 GSTNPVTGVENSFAGVGQKPTIDATLIDTTLNAAEAWEILLOFLSLPLSDSRV
 QKDFSLWTSYGHYGPAPFHEFYENRIANGSVNGVNLNGLINGLIDEAIQ
 APYPEFVNNTGIRAVNETVYMKFAMNGCCQDLISTCKQNRNALADYALCA
 EAINMCRDVEGYYPAGRGVYDIRHPYDDPTPPSYNKFLEADSYMDAIGNYNIT
 QSNDDYVAFQGTDFWPNPFIEDLEILALPVRSYLGADYICNWFQGAQVSLAA
 NYSQAQFRSAGTPLKVGVEYGETREYGNFSFTRVYEAGHEVPYQPIASLQLFNR
 TIFGWDIAEQKKIWFYSYKINGTATATHTQSSVPLPTATSKSSVGNNA"
 sig_peptide 1..54
 /gene="cpi"
 BASE COUNT 548 a 603 c 524 g 570 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.43e-189 Length: 2245
 Score: 2526.00 Matches: 530
 Percent Similarity: 72.54% Conservative: 1
 Best Local Similarity: 72.40% Mismatches: 6
 Query Match: 87.50% Indels: 201
 DB: 8 Gaps: 9
 US-09-712-338-2_COPY_19_555 (1-537) x AF394242 (1-2245)
 QY 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 20
 Db 55 CTCCAGGAAGTACACCGCGTCCGCTCGGTAGAACAGACTACCCAGAACCCACCGGG 114
 QY 21 ValLysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu 40
 Db 115 GTCAAGACTCTTACAAACCGCAACAATGTCCACATCCGGTACAAAGAACCGGGGAGAG 174
 QY 41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60
 Db 175 GCGGTCTGGAGACTACCCGGGTGCAAACTCTCTGGATATGCGACACTCTCC 234
 QY 61 GluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 80
 Db 235 GAGTCCCATACCTCTCTCTGGTCTTCGAAAGCAGACATAACCCAGAACTGCACCAATC 294
 QY 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu----- 98
 Db 295 ACATTGTGTTGAATGGTGGCCCTGGAGCGATTCTTTGATCGGTCTCTCGA-AGTTT 353

QY 99 -----GluLeu 100
 Db 354 GCTTGACGCTGTTACAAAGTANGCTCTTTAGGTCTCGGTTAACCTTGTGTTAGAGTGG 413
 QY 101 GlyProCysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGlu 120
 Db 414 GSCCCTTCGCAATGCAATTCGACTTTTGATGACTACATCAACCCCTCACTCGTGGACGAG 473
 QY 121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGly----- 132
 Db 474 GCTCCCAATTTACTATTCTCTGCCAGCCATT-GGGAGTCGGTANGATTGCCGCCACCT 532
 QY 133 -----ValGlyPheSerTyrSerAsp 139
 Db 533 CCTTACTGGGTCCTCCCATATTGACGAGGTGCTCCCGTAGGCTTTTCATATAATGAT 592
 QY 140 ThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheIleGly 159
 Db 593 ACGGTTGATGGGTCCATTAAACCCCTGTAACCTGGGTCTCGAAATTCGAGCTTTTCAGGA 652
 QY 160 ValGlnGlyArgTyrProThrIleAspAlaThrLeuIle----- 172
 Db 653 GTTCAGGGCCGTACCCCAACCATTCATGCCACTCTGAT-CGGTAACGGTTTCITCCGGTC 711
 QY 173 -----AspThrThrAsnLeuAlaAlaAla 182
 Db 712 CACATGCTCTTATTACTGATACGATACAGATACACCAATCTTCCGCGACAGAGCCGCT 771
 QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
 Db 772 TGGGAGATCCGTCGAAGGATTCCTTAGTGACTACCTAGCTTGGACTCTAGGCTTCAGTCT 831
 QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGly----- 214
 Db 832 AAGGACTTCAGTCTATGACGAGAGCTATGAGG- GTAGTGGTTCCTCCTCGCAATC 890
 QY 215 -----HisTyrGly 217
 Db 891 CTTAAACGGCGTGATGTTGCGCTGATAGTTCCTTCCTTCGAAAGGCATATGTT 950
 QY 218 ProAla-----Phe 220
 Db 951 CCTGCAGTGGTCTTACTTGACCATGTGTCAAAGAAAGCAAGACTAACCAACCTAGTTC 1010
 QY 221 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 240
 Db 1011 TICAATCAITTTACGAGCAGAAAGAGAAATTCGCAACGGTAGTGTTAATGGTGTTCAG 1070
 QY 241 LeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAla--- 259
 Db 1071 CTTAAATTCAACTCTCTGGGAATTAATACGGCATCAICGACGAGGCGATCCAGGT-ATA 1129
 QY 259 ----- 259
 Db 1130 CCGTCGAAACAGCCCATAGAACATCTCCGAAACAAAGCTCTAACGTATGCCCATAGGCC 1189
 QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAla----- 275
 Db 1190 CCTTACTACCTGAATTCGCTGTGACATACCTACCTACCTACCTACCTACCTACCTACCT 1249
 QY 276 -----ValAsnGluThrVa 280
 Db 1250 TTAAGCGATGACATATATCTCTGATACCTATCACCATGAAAGGTCACACGACCGCT 1309
 QY 280 lTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGluAspLeuIleSerTh 300
 Db 1310 CTACAACTCATAGAGATTTCCCAACCAAAATGCGAAATGTTGCCAGGATTGATTTCAC 1369
 QY 300 rCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAs 320
 Db 1370 CTGCAACAGACAGAAACCGCACCGCATAGCTAGTACGGCCTCTGCGCCGAAGCCACCAA 1429

JOURNAL Patent: WO 02086623-A 109 06-SEP-2002;
 DSM N.V. (NL)
 FEATURES Location/Qualifiers
 source 1..1656
 /organism="Aspergillus niger"
 /mol_type="genomic DNA"
 /db_xref="taxon:5061"
 BASE COUNT 411 a 426 c 409 g 410 t
 ORIGIN
 Alignment Scores: Length: 1656
 Pred. No.: 6,37e-175 Matches: 419
 Score: 2338.00 Conservative: 52
 Percent Similarity: 88.87% Mismatches: 59
 Best Local Similarity: 79.06% Indels: 0
 Query Match: 80.98% Gaps: 0
 DB: 6
 US-09-712-338-2_COPY_19_555 (1-537) x AX534871 (1-1656)
 QY 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeuProLysAsnProThrGly 20
 Db 55 ATGCCCGAGATGAATGGTCATCTACGATAAGAAAGGAGTTACCAAAAGGCTCCACATGGC 114
 QY 21 VallysThrLeuThrAlaAsnValThrIleArgTyrLysGluProGlyAlaGlu 40
 Db 115 GTCAATTCGATATAAAACCCCAACAATGTCATCTACGATATAAGAACCAAGGAAACCGAA 174
 QY 41 GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60
 Db 175 GGAATTTGTGAGACAACACCTGGGGTCAATATACATACATCCGATATGCGATCTTCGCCA 234
 QY 61 GluSerHisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIle 80
 Db 235 GAGTCGCATACTTTCTTTTGGTTTTCGAGTACAGCCGTCAGCCGAAATGATCCAGTG 294
 QY 81 ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeu 100
 Db 295 ACTCTGGTGGCTGAATGGTGGCCCTGGAGCGGATCTCTTGATGGCTTTTGAAGAGTGG 354
 QY 101 GlyProCysHisValAsnSerThrPheAspTyrIleAsnProHisSerTyrPasnGlu 120
 Db 355 GGTCCGTGCATACACACAGTACGATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 414
 QY 121 ValSerAsnLeuLeuPheLeuSerGlnProGlyValGlyPheSerTyrSerAspThr 140
 Db 415 GTCACCAATCTCTTTCTTCTGTCAGCCCTCGGCTGGGGTCTCTTACAGTGAACCC 474
 QY 141 ValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyVal 160
 Db 475 GAGCGCGGGTCTCTGAAATCCATTTACTGGAGCCGTCGAGAGCCCTCTCTCTGCTGAGTT 534
 QY 161 GlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrAsnLeuAlaGlu 180
 Db 535 CAGGGTCGATACCCAGTTATTGATGCCACTATCATCGACACGACGATATCGCTGCACGC 594
 QY 181 AlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgVal 200
 Db 595 GCAACCTGGAGAGTCTCTCAGGCTCTCTCAGTGGCTCTCGCAGCTAGATTCGGAATTC 554
 QY 201 GlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhe 220
 Db 655 AAGTCCAGGAGTTCAACCTGTGGACAGAGTACGAGGAGACACTATGACCAAGGCTTC 714
 QY 221 PheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGln 240
 Db 715 TTCAATCATTTCTACGAGCAAAATTCGAAGATCGCTAGCGGGGAGTCAATGCGGCTCAA 774
 QY 241 LeuAsnPheAsnSerLeuGlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaPro 260
 Db 775 CTGAATTTTAACTCCCTCGGATTTATCAACGGGATCATGATGCGCGGATTCAGGACGAC 834
 QY 261 TyrTyrProGluPheAlaValAsnAsnThrTyrTyrIleIleAlaValAsnGluThrVal 280

320 nMetCysArgAspAsnVal----- 326
 1430 CATGTCGAGGACAATGTTGGTACGTCTTTCCCTCCCTTGGAAACAGGTAATGATT 1489
 327 -----GluGlyProTyrTyrAlaPheAlaGlyAr 336
 1490 CAAATAGTGGGACATAACAGCAAAATAGAGGGGCCATACACTACGCCCTTCTGCTGCTG 1549
 336 gGlyValTyrAspIleArgHisProIlyrAsp----- 346
 1550 TGGTGTGTATGATATTTCGGCATCCATATGA-TGTAAGTAGAGCTTTTACTTCTATTATTACA 1608
 346 ----- 346
 1609 CACCGCATGCAATGATGCTGTTTCATCTGTCATGCAATAAAATCCCTTATATGTTATTA 1668
 347 -AspProThrProProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAl 366
 1669 GGACCGGACTCCGCGCAAGTTATTACACAAATTTCTGGCCAGGACACTCTGTCATGGACGC 1728
 366 alileGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnG 386
 1729 TATCGGGGTCAACATCACTACACCCAGTCCCAATATATGAGCTCTACTACGCTTTCCAGCA 1788
 386 nThrGlyAspPheValTrpProAsnPheIleGluAspIleuGluIleLeuAlaLeuPr 406
 1789 AACAGGGCACTTGTCTGGCCCACTTCATCGAAGACCTCGAGGAGATCTCTGCTCTCC 1848
 406 oValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyG 426
 1849 CGTGGGTCTCTCCCTCATCTATGGCGACGCGGATTCATCTGCAACTGGTTCGGCGTCA 1908
 426 nAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrTh 446
 1909 GGGCGTTTCCCTCCCTCGCAACTACTCCCAAGCGCCCGGCTTCCAGAGCGGAGGTACAC 1968
 446 rProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPh 466
 1969 GCCCTGAAAGTCAACGCGCTCGAGTATGGGGAACCTCGCGAGTATGATGTAATTTCTCTT 2028
 466 eThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuG 486
 2029 CACTCGGGTCTATCAGGAGCGCCATGAAGTCCCATACTACAGCCCAATCGCTCCCTGCA 2088
 486 nLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLysIleTrpPr 506
 2089 ATTGTTTAACGGGACTATCTCGGTTGGATATCGAGAGGCGCCAGAGAAGATCTGGCC 2148
 506 oSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSerValProLeuPr 526
 2149 CAGCTACAAGACGAATGAGAGGGCTTACAGCTACGATACACAGTCTGCTCGCTGCGCTGCC 2208
 526 oThrAlaThrSerMetSerSerValGlyMetAla 537
 2209 TACGGCTACACGACATGTCAGTGTGGTATGGCT 2242
 RESULT 3
 AX534871
 LOCUS AX534871 1656 bp DNA linear PAT 22-NOV-2002
 DEFINITION Sequence 109 from Patent WO02068623.
 ACCESSION AX534871
 VERSION AX534871.1 GI:25261462
 KEYWORDS
 SOURCE Aspergillus niger
 ORGANISM Aspergillus niger
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes.
 Eurotiiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
 1
 Edens L., dijk Van A.A., Krubasik P., Albermann K., Stock A.,
 Kimpel E., Klugbauer S., Wagner C., Fritz A., gustadt Von W.,
 Heinrich O., Maier D., Spreafico F., Folkers U., Hopper S.,
 Kemmer W., Tan P., Stiebler J. and Albarg R.
 Novel genes encoding novel proteolytic enzymes

```

Db      835  TACTACCGAGACTTGGCGTTAAATAATACATATGCAATCAAGCTGTCAATGACACAGTG 894
Qy      281  TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThr 300
Db      895  TACAACATATATGAGTTCGCGAACACAGATGCCAATGATGATGCCAGGATGAGTTCGTCG 954
Qy      301  CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 320
Db      955  IGTAATATGACCAATAGACCTGCTTCGTATATGCTATATGCTATGCTACAGAGCCCAAT 1014
Qy      321  MetCysArgAspAsnValGluGlyProTyrThrAlaPheAlaGlyArgGlyValTyrAsp 340
Db      1015  ATGTCAGAGGACAAATGTCGAGGCGCTTACTACCAAGTTTGGCGCGCTGTATGAT 1074
Qy      341  IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 360
Db      1075  ATTCGGGACCCCTACATATGATGACCCCGCGCTTCTTGTGCTACTTACCTCAAGAA 1134
Qy      361  AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 380
Db      1135  GACTCAGTCATGATGATGTCGCGGTGAGCAITTAACCTACCGAGTCCAGCGCGAGTA 1194
Qy      381  TyrTyrAlaPheGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGlu 400
Db      1195  TATATGCAITTCAGACAGCGCGGACTTGTATGGCGGAATTCATTGAGGACCTCGAA 1254
Qy      401  GluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCys 420
Db      1255  GAGATCCTCAACTCCCGTACCGGTCGTTGATCTACCGGCGATGCGGACTATATCTGI 1314
Qy      421  AsnTrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPhe 440
Db      1315  AACTGTTCGCGGTGAGGCGCATCTCAGTTCAGTTAACTACCCCGATGCGAGTCTCAG 1374
Qy      441  ArgSerAlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGlu 460
Db      1375  CGTGCAGCGGATACACACCATGACAGATAGATGGGTGCAATACGGTGAGACTCGCGAG 1434
Qy      461  TyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGln 480
Db      1435  TATGCAACTTTCGTTCACCGCGTATATACGAGTGGGCGAGGTTCCATCTACTATCAA 1494
Qy      481  ProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGly 500
Db      1495  CCGATCGCAGCGTTCAGCTGTCAACCGTACTTTATTTGGATGGGATATTCAGCGGGT 1554
Qy      501  GlnLysLysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGln 520
Db      1555  ACAACTCAGATTGGCCCGGAATATAGCAACCGGACATCGCAGGCTACACACAGGAG 1614
Qy      521  SerSerValProLeuProThrAlaThrSer 530
Db      1615  TCGTTCGTCGCCATGTCACCGGCGTCAAGT 1644

```

RESULT 4

AX534814

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AX534814
Sequence 52 from Patent WO02068623.
PAT 22-NOV-2002

AX534814.1 GI:25261343

Aspergillus niger
Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1

Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A.,
Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,
Heinrich, O., Maier, D., Spreafico, F., Folkers, J., Hopper, S.,
Kemmer, W., Tan, P., Stiebler, J. and Albarg, R.
Novel genes encoding novel proteolytic enzymes

Patent: WO 02068623-A 52 06-SEP-2002;

FEATURES

source Location/Qualifiers
1..3150
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"

BASE COUNT 801 a 778 c 738 g 833 t
ORIGIN

Alignment Scores:

Pred. No.: 1.32e-146 Length: 3150
Score: 1983.00 Matches: 412
Percent Similarity: 65.07% Conservative: 50
Best Local Similarity: 58.03% Mismatches: 68
Query Match: 68.69% Indels: 185
DB: 6 Gaps: 8

US-09-712-338-2_COPY_19_555 (1-537) x AX534814 (1-3150)

```

Qy      1  LeuProGlySerThrProAlaSerValGlyArgArgGlnLeuProLysAsnProThrGly 20
Db      629  ATGCCCGAGAAATGAATGTCATCTACGATAAGAAGGCAGTTACCAAAAGCGTCCACTGCG 688
Qy      21  ValLysThrLeuThrAlaAsnAsnValThrIleArgTyrLysGluProGlyAlaGlu 40
Db      689  GTCAAATCGATAAAACCCCAACAATGTCATCTAGGTATAAGAACACCGAACCGAA 748
Qy      41  GlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThrSerPro 60
Db      749  GGAATTTGTGAGACACACACCTGGGTCAATCACTCCGGATATGTCGATCTTCGCCA 808
Qy      61  GluSerHisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIle 80
Db      809  GAGTCGCATACITTTTIGTTTTCGAGTCACGCGGACCGGACCGGACCGGACCGGAC 868
Qy      81  ThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu 98
Db      869  ACITCTGGCTGAATGGTGGCCCTGGAGCGATTCCTTGATTTGGCTTTTGA-AGGTTG 927
Qy      98  -----
Db      928  GCCAAATATCTCTGACGGGAAAGATAAATTCAGTTCATGCTCTGACGCTTCACAACA 987
Qy      99  GluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrp 118
Db      988  GAGTTGGTCCGTGTCACATCACACAGAGTAGGAATCAATCATCATCATCTACTCTCTG 1047
Qy      119  AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyVal 133
Db      1048  AACGAGTCAACCAATCTCTTTTCTGTCTACGCCCTCGGTGGGTATGGGAATATGTC 1107
Qy      134  -----
Db      1108  TGCCITTCATACATCTCTGAGTACATTCCTTACGGTCTTATCIGCGAAGGTTCTCTACAG 1167
Qy      138  rAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAl 158
Db      1168  TGAACCGAGCGCGGCTCTTGAATCATTTACTGGAGCGCTCGAGAACCGCTCTCTTGC 1227
Qy      158  aGlyValGlnGlyArgTyrProThrIleAspAla 169
Db      1228  TGGAGTTCAGGGTCGATACCCAGTATGATGCCATATCATCGGTAAGTTCGCGTTT 1287
Qy      170  -----
Db      1288  GACTCTCCTACCTAGATTCCTCAATGTCCTACTTACACACCGACCGATATCCTCTGC 1346
Qy      179  aGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerAr 199
Db      1347  ACAGCGCAACTGGGAGTGTCTTACGGGCTCTCTCTAGTGGCTCTGACAGTATGCCA 1406
Qy      199  gValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGly-GlyHisLysTyrGly 217

```

Db	1407	AGTCAAGTCCCAAGGAGTTCAACCTGTGGACAGAGAGTTACGGAGGGTGAGTGCACACTTTC	1466
QY	217	-----	217
Db	1467	ATACCAGACCGACGTAAGCTGACTGATCAAGACACTATGGACCGGAGTGTCTT	1526
QY	218	-----	218
Db	1527	TTCTGGTGCACACATATGATGACTAATGACCGAAGTCTTCAATCATTTCTACGAGCAAA	1586
QY	228	snGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyL	248
Db	1587	AUTCAGAGATCGTAGCGGGAAGTCAATGGGCTCCCAACTGAATTTAACTCCCTCGGA	1646
QY	248	leIleAsnGlyIleAspGluAlaIleGluAlaPro-----	260
Db	1647	TTATCAACGGCATCATGATGCGGCGGATTCAGGTACT-TAGAAATGCAGCTCCGCGAGAG	1705
QY	261	-----	261
Db	1706	GCTGCGCCTAGNAGGACATCGCTAAAGTAATTAATAGGACAGACTACTACGAGACTTTG	1765
QY	266	laValAsnAsnThrTyrGlyIleLysAla-----	275
Db	1766	CGGTTAATAATACATATGGAATCAAAAGCTGTAAGTTTAATACACGTCATCGTGGATTT	1825
QY	276	-----	276
Db	1826	AGATCAACCGTGCTCATGCTTCTAGGTCAATGACACAGTGTACACATATATGAGATTC	1885
QY	287	AlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArg	306
Db	1886	GCCACACAGATGCCAATGATGCCAGGATCAGGTTGCTCGTGTAAATTCACCAATAGG	1945
QY	307	ThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnVal	326
Db	1946	ACCTCGCTTCTGATTATGCTATATGTTACAGAGCAACCAATATGTCAGGGACAATGT-	2004
QY	326	-----	326
Db	2005	CGGTGAGTGTCTTCTGTCGAGGGGTGCAATGATGAGGACTTTGCTAAGCTG	2064
QY	327	-----	327
Db	2065	TCATGTACAGAGGGCCCTTACTACCGATTTGGCGCGTGGCGGTATGATATCGGCAC	2124
QY	344	ProTyrAsp-----	346
Db	2125	CCCTACAA-TGTAAGTGGCAAGGATAAGGATTGTACTTCCGAACAGGACACTGCTCAT	2183
QY	347	-----	347
Db	2184	ATGTCAACGTAGGACCGACCCCGCTCTACTTTGTTGACTACCTCAAGAAAGACTCA	2243
QY	363	ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyr	382
Db	2244	GTCATGATGCTATCGCGTGGACATTAACATACACCGAGTCCAGCGGCGAAGATATAT	2303
QY	383	AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIle	402
Db	2304	GCAATCCAGCAGACCGGCGACTTTGATGGCGGAATTTCAATTGAGGACCTCGAGAGATC	2363
QY	403	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	422
Db	2364	CTCCNACTCCCGGTACGGTGTGCTGATCTACGGCGATGCCGACTATATCTGTAAGTGG	2423
QY	423	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaGlnPheArgSer	442
Db	2424	TTGCGCGGTGAGGCCATCTCACTCGCAGTTAACTACCCCATCGCAGCTCAGTTCGGTGCA	2483
QY	443	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly	462
Db	2484	CGGGATACACACCATGACAGTAGATGGGTTCGAATACGGGTGAGACTCGCGGAGTATGAC	2543
QY	463	AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle	482
Db	2544	AACTTTTCTCCCGCGGTATATCATAGCTGGCGACGAGTTCCATACACTATCAACCGATC	2603
QY	483	AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTyrAspIleAlaGluGlyGlnLys	502
Db	2604	GCAGCTTTCAGCTGTTCACCGTACTTATTTGATGGATGATATTCGCGGGGTACAACT	2663
QY	503	LysIleTrpProSerTyrLysThrAsuGlyThrAlaThrAlaThrHisThrGlnSerSer	522
Db	2664	CAGATTGGCCCGAATATAGCAACCAAGGACATCGCAGGTACACACACGAGGTGCTTC	2723
QY	523	ValProLeuProThrAlaThrSer	530
Db	2724	GTGCCACTGTCCACGGGCTCGAGT	2747
RESULT	5		
AX534872			
LOCUS		1872 bp	DNA
DEFINITION		Sequence 110 from Patent WO02068623.	
ACCESSION		AX534872	
VERSION		AX534872.1	GI:25261464
KEYWORDS			
SOURCE		Aspergillus niger	
ORGANISM		Aspergillus niger	
REFERENCE			
AUTHORS		Edens, L., Dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A., Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W., Heinrich, O., Mater, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albang, R.	
TITLE		Novel genes encoding novel proteolytic enzymes	
JOURNAL		Patent: WO 02068623-A 110 06-SEP-2002;	
FEATURES		DSM N.V. (NL)	
source		1. 1872	
		/organism="Aspergillus niger"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:5061"	
BASE COUNT		402 a 434 c 570 g 466 t	
ORIGIN			
Alignment Scores:			
Pred. NO.:		6.68e-68	Length: 1872
Score:		980.50	Matches: 216
Percent Similarity:		56.23%	Conservative: 91
Best Local Similarity:		39.56%	Mismatches: 178
Query Match:		33.96%	Indels: 61
DB:		6	Gaps: 19
US-09-712-338-2_COPY_19_555 (1-537) x AX534872 (1-1872)			
QY	26	ThrAlaAsnAsnValThrIleArgTyrLysGlu-----	ProGlyAlaGlu----- 40
Db	94	ACGCGTGAAGATCTCACCGTTATTCATTCGGAGATATTCCTCGGTCGAGGATCTCCAT	153
QY	41	-----	41
Db	154	AAGCAACCCCTCGGATCTGCACACACCCCTCCACCCCTCCACCCCTCCACCCCTCCAC	213
QY	57	AspThrSerProGluSer-----	62
Db	214	CACCTCCCGCCACACACCGCTTACCAATCTCTCCATCCAGGAATCAGCATCTCCGAACA	273
QY	63	-----	63
Db	274	TACCTATCAATACCTTTTCTTCTGTTACCTTCTCTCCGCCATCACCACAAATGATACA	333
QY	78	AlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe	97
Db	334	TCCCGGATACACCATCTGGATGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	393

QY	98	Gl uGluLeuGlyProCysHisValAsnSerThrPheAspAsp--TyrIleAsnProHis	111
Db	394	CAAGAGAACGGGCCCATGTACTGTGAATACGGGACICGAATCCACGGCCCTATATCCCTGG	453
QY	117	SerTirAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer	136
Db	454	TCGTGAATAGTACGTCGATGTGTATATGATGACGACGGTGCAGCGGATTAGT	513
QY	137	TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer	156
Db	514	TATGATGTCTTACGAATAGGGACGTATAGT-----ITGAATGAGACG	555
QY	157	PheAlaGlyValGlnGlyArgTyrProThrIleAsp-----AlaThrLeuIleAspThr	174
Db	556	TTT-----TTGGTGGGACGTTGCCGAGTCAGGATGTCATGGGACGGTGAATGGGACG	609
QY	175	ThrAsnLeuAlaAlaGluAlaAlaThrPgluIleLeuGlnGlyLeuSerGly-----	192
Db	610	GTTAAT-----GGGGGAAGGGCGCTTTGGCTGGCAGGTTTGGTGGGTGAATTCCT	666
QY	193	-----LeuProSerLeuAspSerArgValGlnSerLysAsp---PheSerLeuThrThr	209
Db	667	CAATATGTTCTCTGTGTGACGGGAATGGTGGTGTATGACACGGGTGAGTATATGGACG	726
QY	210	GluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGluAsnGlu	229
Db	727	CAGTCATATGGGGACGGTATGGACCGGCATACACGGCGCTCTTTCAGGAGATCAATGAG	785
QY	230	ArgIleAlaAsnGlySerValAsn---GlyValGlnLeuAsnPheAsnSerLeuGlyIle	248
Db	787	AGGATTGAGAGTGGGGAGGTAAAGCACCGGAAGAGATCCATTTCGATACGCTGGGCATT	846
QY	249	IleAsnGlyIleIleAspGluAlaIleLeuAlaProTyrTyrProGluPheAlaValAsn	268
Db	847	ATCAATGGTGTGTCGATTACTCTGTCGAGTCCCTTCGTCCCTGACGAGCGGTATAAC	906
QY	269	AsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsn	288
Db	907	AATACGATGGGATCGAGGAATCAATCGCACGCTCTACGACCGGGCTATGGATAGTTGG	966
QY	289	GlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAla	308
Db	967	AGCAAGCCTGGCGGGTTCAGGGATATGATCATCGAGTGCAGTGCAGTGCAGTGCAGT	1020
QY	309	LeuAlaAsp--TyrAlaLeuCysAlaGluAlaPheAsnMetCysArgAspAsnValGlu	327
Db	1021	CTCGGAGATCCCTCATCTCGGAGGAGCGTCGACACTACTGTCGGGGAGATCAAG	1080
QY	328	GlyProTyrThrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAsp	347
Db	1081	AGCCTGTATACGAATACCTCCGGGAGGATACTACGACATAGCGCATTCACGCCCGAT	1140
QY	348	ProThrProSerTyrThrAsnLysPheLeuAlaLysAspSerValMetAspAlaIle	367
Db	1141	GCAGCTCTCGTGGCTTACTTCGTGGGTTCTTGAATGCCCATGGGTGCAAAAGCACTT	1200
QY	368	GlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnThr	387
Db	1201	GGGCTCCGGTGACATATACCATGTCTGTCAGAGGAGTGGGGAACAGTTCGCTCGACG	1260
QY	388	GlyAspPheValTrpPro-----AsnPheIleGluAspLeuGluIle	402
Db	1261	GGCGATTAT-----CCGCGAAATGATCCCGCGGAATGATCGGGGATATGGAACTTG	1314
QY	403	LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp	422
Db	1315	CTTGACTCCGGTCAAGGTGGCTATGGTATATATGGGACCGGACACTGCTTCTCCGTGG	1374
QY	423	PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerClnAlaAlaGlnPheArgSer	442
Db	1375	CGCGCGGGGAACATGTACCTTCCTGGTGGAGTACGAGGATCGGAGAGATTCCTGCT	1434

QY	443	AlaGlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGlu	460
Db	1435	 TTTTTT GGTGGTATGCCGAAGTCACAGAAGTCA--TCCTACGTTGGGGTCTAGTAGGCAG	1491
QY	461	TyrgLyAsnPheSerPheThrArgValTyrgLuAlaGlyHisGluValProTyrrGln	480
Db	1492	 TTTTTT TATGGAACTTCTCGTCCACGCTGCTTTTCAGCGGGCCCATGAGGTGCCATTATCAG	1551
QY	481	ProIleAlaSerLeuGluLeuPheAsnAArgHrIlePheGlyIrpAspIleAlaGluGly	500
Db	1552	 TTTTTT CCCCAACGCGCATGATGATTTTAATGCCCTCAGTTAATCGGATATTCGACACGGGA	1611
QY	501	GlnLysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThraLthrHis	518
Db	1612	 TTTTTT GCAFTTCTTCGGAGCAGAAATCAGAGCTATGGAGCGGAGCGTGTCAACGTGGCAT	1671
QY	519	ThrGlnSerSerValPro 524	
Db	1672	 ATCAAAGCAAGTGCCG 1689	
RESULT 6			
AX534815			
LOCUS	AX534815	3221 bp DNA linear	PAT 22-NOV-2002
DEFINITION	Sequence 53 from Patent WO02068623.		
ACCESSION	AX534815		
VERSION	AX534815.1 GI:25261346		
KEYWORDS			
SOURCE	Aspergillus niger		
ORGANISM	Aspergillus niger		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
AUTHORS	Edens,L., dijk Van.A.A., Krubasik,P., Albermann,K., Stock,A., Kimpel,E., Klugbauer,S., Wagner,C., Fritz,A., gusted Von.W., Heinrich,O., Mater,D., Spreafico,F., Folkers,U., Hopper,S., Kemmer,W., Tan,P., Stiebler,J. and Albang,R. Novel genes encoding novel proteolytic enzymes Patent: WO 02068623-A 53 06-SEP-2002; DSM N.V. (NL)		
TITLE			
JOURNAL			
FEATURES	Location/Qualifiers		
source	1..3221		
	/organism="Aspergillus niger"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:5061"		
BASE COUNT	739 a 773 c 869 g 840 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	2,47e-67	Length:	3221
Score:	977.00	Matches:	213
Percent Similarity:	55.60%	Conservative:	90
Best Local Similarity:	39.08%	Mismatches:	172
Query Match:	33.84%	Indels:	70
DE:	6	Gaps:	17
US-09-712-338-2_COPY_19_555 (1-537) x AX534815 (1-3221)			
QY	41	GlyValcysGluThrThrProGlyValLysSerGlyTyrValAspThrSerPro	60
Db	918	GGCATCTGCACCACACCCCCTCCACCCCGACTACTCGGCTACATCCACCTCCGCCA	977
QY	61	GlusEr-----His 63	
Db	978	CACACCCCTTACCATCTCTCCATCCAGCAATCAGCATCTCGCACACCATACCTTACAT	1037
QY	64	ThrPhePheTrpPheGluAlaArg-----HisAsnProGluThrAlaProIleThr	81
Db	1038	ACCTTTTCTGTGTPACTTCTCTCCGCCCATCACCACACAATGATACATCCCACTACC	1097
QY	82	LeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGlnLeuGly	101
Db	1098	AUCTGATGAACGGCGGGCGCGGAICTCCATGATTGGCTATTTCAGAGAACGGG	1157

QY 102 ProCysHisValAsnSerThrPheAspAsp---TyrIleAsnProHisSerTrpAsnGlu 120
 DB 1158 CCATGCTACTGTGATACGACACGATCCACGGCCCTATATCCCTGGTCTGGTGAATGAG 1217
 QY 121 ValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThr 140
 DB 1218 TAGTCGATATGTTGATATGACGACGGGTGACACGGGATTTAGTTATGATGATG 1277
 QY 141 ValAspGlySerIleAsnProValThr----- 149
 DB 1278 AGGAATGGACGCTAGATTTGGTGTAGTGGGAGATAGATTTAGTATTACTGAAGTGAG 1337
 QY 150 -----GlyValValGlnAsnSerPheAlaGlyValGlnGlyArgTyrProThrIle 167
 DB 1338 AGGATGGAGTAGACAGATGACACCTTT-----TTGGTGGGACGTTGCCGAGTCAG 1391
 QY 168 Asp-----AlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 185
 DB 1392 GATGTCATGGGACGCTGAATGGAGCGTTAAT---GGGGAAAGGGCGCTTTGGGTGGG 1448
 QY 186 LeuGlnGlyPheLeuSerGly-----LeuProSerLeuAspSerArgValGln 201
 DB 1449 TTGCAGGTTGGTGGGTGAATTCCTGTAATATGTTCTCTGTTCAGGGGAATGGTGGI 1508
 QY 202 SerLysAsp---PheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhe 220
 DB 1509 GGTGATGACAGGCTGAGTATATGACGAGTATATGGGACGATGACGCGGCATAC 1568
 QY 221 PheAsnHisPheTyrGlnGlnAsnGluArgIleAlaAsnGlySerValAsn---GlyVal 239
 DB 1569 ACGGCGCTCTTTCAGGAGATGAATGAGAGGATTCAGAGTGGGAGGTAAAGCAGGGAAG 1628
 QY 240 GlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAla 259
 DB 1629 AAGATCATTTGGATGCGGGCATATATCAATGGGTGTGTGATTTACTCGTCGACGTC 1688
 QY 260 ProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThr 279
 DB 1689 CTTCTGCTCCCTGACGAGCGGTATACAAATACGATGGATCGAGGGAATCAATCGCACG 1748
 QY 280 ValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSer 299
 DB 1749 CTCTACGACCGGCTATGGATAGTTGGAGCAAGCCIGCGGGTGTGAGGATATGATCATC 1808
 QY 300 ThrCysLysGlnThr-----AsnArgThrAla 308
 DB 1809 GAGTGTCCGATGTCGCGAGCTCGGAGATCCCTCATGATGCGGCAATGACAGCGTA 1858
 QY 309 LeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGly 328
 DB 1869 -----AATACATCTCGAGGAGGCGTGGACTACTGTTCCGCGGAGATCAAGAGC 1919
 QY 329 ProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro 348
 DB 1920 CTGTATACGAATACCTCCGGCGAGGATACACATAGCATGCGCATTCACGCCGGATGCA 1979
 QY 349 ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGly 368
 DB 1980 GCTCTCGTGTCTACTTCGTCGCGTCTTGTGAATCGCCCATGGGTGCAAAAGCACTGGG 2039
 QY 369 ValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThrGly 388
 DB 2040 GTCCCGGTGAATATACCATGTCGTCAGAGCGATGGGGAAACAGTTTCGCTCGACGSGC 2099
 QY 389 AspPheValTrpPro-----AsnPheIleGlnAspLeuGluGlnIleLeu 403
 DB 2100 GATTAT-----CCGCGAAATGATCCCGCGGATGATGCGGGATATGGATCTGCTT 2153
 QY 404 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 423
 DB 2154 GACTCCGGTGTCAAGGTGGCTATGATATGCGGACCGGAGTATGCTGTGTCGCGGCGC 2213
 QY 424 GlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 443

DB 2214 GGCGGGAGAGATGACGCTGCTGGTGGAGTACAGAGATCGGAGAGATTCCTGCGCT 2273
 QY 444 GlyTyrThrProLeuLysValAsnGlyValGluTyr-----GlyGluThrArgGluTyr 461
 DB 2274 GGATATCCCGAAGTGGAGACCAAGTCA---TCCTACGTGGGGTCTAGTAGGCAAGTAT 2330
 QY 462 GlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnPro 481
 DB 2331 GGGAACCTCTCGTTCACGCGTCTTTTCAGGGGGCCATGAGGTGCCATTTATCAGCC 2390
 QY 482 IleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGln 501
 DB 2391 GAAACGCGGATGAGATTTTAAATCGCGCTCAGTTAATGGGATATTCGACGGGAGC 2450
 QY 502 LysLysIleTrp-----ProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThr 519
 DB 2451 ATTTCCTGGAGCAGATCAGAGCTATGGGACGAGGACCGTCGTCAACGTGGCAATC 2510
 QY 520 GlnSerSerValPro 524
 DB 2511 AAAAACGAAGTCCG 2525
 RESULT 7
 AX534866
 LOCUS
 DEFINITION Sequence 104 from Patent WO02068623.
 ACCESSION AX534866
 VERSION AX534866.1 GI:25261452
 KEYWORDS
 SOURCE
 ORGANISM
 Aspergillus niger
 Aspergillus niger
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 REFERENCE
 AUTHORS
 Edens, L., dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A.,
 Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustedt Von, W.,
 Heinrich, O., Mater, D., Spreafico, F., Folkers, O., Hopper, S.,
 Kemmer, W., Tan, P., Stiebler, J. and Albang, R.
 Novel genes encoding novel proteolytic enzymes
 Patent: WO 02068623-A 104 06-SEP-2002;
 DSM N.V. (NL)
 TITLE
 JOURNAL
 FEATURES
 Location/Qualifiers
 1..1665
 /organism="Aspergillus niger"
 /mol_type="genomic DNA"
 /db_xref="taxon:5061"
 BASE COUNT 374 a 455 c 424 g 412 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.48e-65 Length: 1665
 Score: 941.00 Matches: 209
 Percent Similarity: 51.75% Conservative: 72
 Best local Similarity: 38.49% Mismatches: 172
 Query Match: 32.59% Indels: 90
 DB: 6 Gaps: 13
 US-09-712-338-2_COPY_19_555 (1-537) x AX534866 (1-1665)
 QY 13 GlnLeuProLysAsnProThrGlyValLysThrLeuThrAla-----AsnAsnVal 30
 DB 49 CAATTCCTCCGCGGAGGCAATCACTGTGCTCAAGTCCAGTCCAGTCAGAGATGIG 108
 QY 31 ThrIleArgTyrLysGluProGlyAlaGluGlyCysGluThrThrProGlyValLys 50
 DB 109 ACTATTTCTTTCAGAGAGCTT-----GGAATTCGAAACTACCGCGGTGCGGA 159
 QY 51 SerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPheGlu 70
 DB 160 TCTTATTCGGGTATGTACACTTCCCGCCCGCTCAACAGCACTCTTTTGGTTTCGAA 219
 QY 71 AlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySer 90


```

Db      220  GCGCGAAGATCCACGAAATGCGCTCGGCATCTGCTCAATGCGGTGGGTGGC 279
QY      91  AspSerLeuIleGlyPheGluGluLeuGlyProCysHisValAsnSer----ThrPhe 109
Db      280  TCGTCGCTCATGGGCTCTTGAAGAATTAGGTCCTGTGTCATCATCAGACTCCAG 339
QY      110  AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln 129
Db      340  ACCACAGTCTCAATCTTGGAGTTGGAACAATGAAGTCAATCTTATCTTCCCTTGACCAG 399
QY      130  ProLeuGlyValGlyPheSerTrpAspThrValAspGlySerIleAsnProValThr 149
Db      400  CCAACTCAAGTCGGGTTCATACGATGTCACCAACAAATGGCACTTGACA----- 450
QY      150  GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla 169
Db      450  ----- 450
QY      170  ThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPhe 189
Db      451  -----GCTAATGGGACTGCATTCGCGGCTCAGCTCATGCAATTCGCGCAACACCTGG 504
QY      190  LeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerLysAspPheSer 206
Db      505  TTTTTCGAGTCTCCACACTACAAAGCCAAACGATGATCGTGTG-----ACT 549
QY      207  LeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGlu 226
Db      550  CFTCGGCTGAAAGTACGAGGACCATATGTCAGGCACTTTCGGTTCCTCCAAACAS 609
QY      227  GlnAsnGluArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPheAsn 244
Db      610  CAGAAATGACAAATCGCAGAGGGGACTGCAGAAAGCGGTGCACAGTATTGTCATCTGCAC 569
QY      245  SerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTrpProGlu 264
Db      670  ACCCTTGCAATGTGAACGGTGTGATGATATGTTGATCCCAAGAGAGCGCTTACATTACT 729
QY      265  PheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsnTyrMet 284
Db      730  TGGCCA-----TACAATAACGTA 747
QY      285  LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db      748  AGGCTCGCCCT-----TCITCATTC 768
QY      305  AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaAlaThrAsnMetCysArgAsp 324
Db      769  AACTCGGAGGCTTCGGGATCAGGCCCTCGCCTCGCAAGCGCTTTGAAAGAACGGCAT 828
QY      325  -----AsnVal----- 326
Db      829  TCGGCTTGCCCTCACTCAGGAAGAATATCTCTGAATTCGGGAGGCTTCGACTAGAA 888
QY      327  -----GluGlyPro-----TyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArg 342
Db      889  TGGGAGATGCGCCCATCACCCTACTACACACCTTCAATCGCGGTGTGTAGACATCGCC 948
QY      343  HisProTyrAspAspProThrProProSerTyrTrpAsnLysPheLeuAlaLysAspSer 362
Db      949  CATCTTAAGACGACCCATTCCTCGCAAGCACATGCTCGGATATTTCAGCAGGAGTCC 1008
QY      363  ValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyr 382
Db      1009  GTCTCTCGGCTCTCTGGGTACCAAGTCAATTCATCGCTTCGAGTGGCGIGTCACACA 1068
QY      383  AlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluGluIle 402
Db      1069  CAGTTCAATAAAACCTTTGATATGCTCCACGGCGCTTCCTGGATGCAATTTGGTACTCTC 1128
QY      403  LeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrp 422

```

```

Db      1129  CTCGACAGTGGTGTAAAGTACACATGATGACGGAGATCGTGATATACGCTGCAATGG 1188
QY      423  PheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSer 442
Db      1189  GTCGGGGCGGAAAAAGCCAGCCITGCAGTTCCTGATTCCTGATTCCTGATTCCTGATTC 1248
QY      443  AlaGlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462
Db      1249  ACCGGATATCTCCCATCTTACGCCCGCGGATCAGCGGCATGACCCCGACGCTGGGC 1308
QY      463  AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482
Db      1309  AACGACGCTCTACTCGGTCTTCCAGCGCGGATGAGGTCCCTCCCTACCGACGCTGC 1368
QY      483  AlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAlaGluGlyGlnLys 502
Db      1369  GCGGCGTATGAGATCTTCATGCGCGGCACATTCAACAAGATATCCCTACTGCCCTTTC 1428
QY      503  LysIleTrpProSerTyrLysThrAsnGlyThrAlaThrAlaThrHisThrGlnSerSer 522
Db      1429  GCTGTGTGATGACGAATTCAGTGGTGGACCTTAGGATACGTGGCATATCAAGAATATC 1488
QY      523  ValProLeu 525
Db      1489  CCTCCTATT 1497

RESULT 8
AX534824
LOCUS      1581 bp      DNA      Linear      PAT 22-NOV-2002
DEFINITION      Sequence 62 from Patent WO02068623.
ACCESSION      AX534824
VERSION      AX534824.1      GI:25261366
KEYWORDS
SOURCE
ORGANISM      Aspergillus niger
REFERENCE
AUTHORS      Edens, L., dijk Van, A. A., Krubasik, P., Albermann, K., Stock, A.,
Kimpel, E., Klugbauer, S., Wagner, C., Fritz, A., gustadt Von, W.,
Heinrich, O., Maier, D., Spreafico, F., Folkers, J., Hopper, S.,
Kemmer, W., Tan, P., Stiebler, J. and Albang, R.
TITLE      Novel genes encoding novel proteolytic enzymes
JOURNAL      Patent: WO 02068623-A 62 06-SEP-2002;
FEATURES
            Location/Qualifiers
            source
            1..1581
            /organism="Aspergillus niger"
            /mol_type="genomic DNA"
            /db_xref="taxon:5061"

BASE COUNT      343 a  473 c  423 g  342 t
ORIGIN

Alignment Scores:
Pred. No.:      5,98e-63      Length:      1581
Score:          916.50      Matches:      212
Percent Similarity: 51.97%      Conservative: 78
Best Local Similarity: 37.99%      Mismatches: 187
Query Match:      31.75%      Indels:      81
DB:              6      Gaps:      13

US-09-712-338-2_COPY_19_555 (1-537) x AX534824 (1-1581)

QY      9  ValGlyArgArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrAlaAsn 28
Db      46  GTCTCCCGGGCCAGTTTGTGCTCGCCCGCGATCTCATCCACCAAGGATATCTC 105
QY      29  AsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGly 48
Db      106  GACATCCCGCTCCGCTACAAACAGTCC---CCCACCGSCATTTGTGAGACTGATCCCACT 162
QY      49  ValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPheTrpPhe 68

```

```

Db      163  GTCAGAGCTTCCTCCGGTTCAGTCGATCGCTGAGCATGAGCACATCTTCCTTCCTGGTTC 222
QY      69  PheGluAlaArgHis---AsnProGluThrAlaProIleThrLeuTrpLeuAsnGly--- 86
Db      223  TTCAGGGCGCAACCAAGATCCACACGAGGCTCCCTTGACCGTCTGGATCAATGAGGC 282
QY      87  -----GlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
Db      283  ATGCTGACCCCGGTCCTGGTTCCTCCCATGATCGGCTTGTTCCAGAGACACGCCCA 342
QY      103  CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerIleAsnGluValSer 122
Db      343  TCCGGCATGACGCCAATGGCTCCGCTCTACAACACCCCTACTCTCGAACAACGCCAGC 402
QY      123  AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
Db      403  AACATGCTCTACATCGACGAGCCGCTGACAGCCGCTCTCTCTACAGC----- 450
QY      143  GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db      451  -----ATTCCGGTTCCTCCGGCTATGTGGATCTTCCACAGACAATGGTTTATGGGC 501
QY      163  ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaAla 182
Db      502  GCATTTCTCT----- 510
QY      183  TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db      511  -----CAGTACTCGCGC 522
QY      203  LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 222
Db      523  GAAACCTTCACCTTCACACGAGAGTATGCGGCCCTACGGGCCCTCTTCACAGAG 582
QY      223  HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db      583  TACATCGAGGACAGAACCCCATCTCCAGCGGAGGC-----AAGAAATCCAA 633
QY      243  PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyr 262
Db      634  CTGGCAGTGTGATGATCGCAATGGCTGTATGACCCGATATTCATACACGGCCTAC 693
QY      263  ProGluPheAlaVal-----AsnAsnThrTyrGlyIleLysAlaValAsnGluThrVal 280
Db      694  TACAACCTTACGTATATCCGSCAACACATACGACTACCTGCCATTCAACAAGTCCATC 753
QY      281  TyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspIleSerThr 300
Db      754  AGCTCGCTGATGTACAAACACCTCTATGCGCCCGGAAACTGCCTCGACGCTCTACGAC 813
QY      301  CysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsn 320
Db      814  TGC-----GCCGCCGAGGCATCGACGAGATCTGCAGCACTGCCGACGAT 858
QY      321  MetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAsp 340
Db      859  TTTTGGCCCAACGAGGTGCAAAACGCTACGACATTTACTCCGCTCGGATGAGTATGAC 918
QY      341  IleArgHisProTyrAspAspProThrProProSerTyrTyrAsnLysPheLeuAlaLys 360
Db      919  TTTTCGTGAATCACTACTCCGACCCGCTCCCTACGAGTCTTACGTTGACTACCTCCGACAAA 978
QY      361  AspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGlnSerAsnAspVal 380
Db      979  GCGTCCGTGAGCGCCGATCGGGCATACATCAATTACACGAGAGCAACACGCTGTT 1038
QY      381  TyrTyrAlaPheGlnGlnThrGly---AspPheValIleProAsnPheIleGluAspLeu 399
Db      1039  GGACTCGCTTTTCTCCACCGGTGACGAGCGGCGACTCATGACACACCATCCAGAGTGT 1098
QY      400  GluGluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIle 419
Db      1099  GGCAAGTGTCTCAACAGGGGTGTACCGGTGTGTACGCGGGGAGTGGCGACTATAAC 1158

```

```

QY      420  CysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln 439
Db      1159  TCAACTGGCTGGTGGGAGCCGCTGCTTGCAGGTC-----AAGGCCGCCAAC 1209
QY      440  PheArgSerAlaGlyTyrThrProLeuLysVal---AsnGlyValGluTyrGlyGluThr 458
Db      1210  TTCAGTAGTCGGGTACACCAACATGTCACTCGGATGGAGTACACACCGCGCAGG 1269
QY      459  ArgGluTyrGlyAsnPheSerPheArgValTyrGluAlaGlyHisGluValProTyr 478
Db      1270  CCCCAGGGGGCAATTGCCCTTGTGCGAGTGTATGACAGTGGACATGAGGTTCCTTC 1329
QY      479  TyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIlePheGlyTrpAspIleAla 498
Db      1330  TATCAACCTTCTGCTGGCTGGAGATGTTGAGCGGCTCATTTGGCGCAAGATGTGGG 1389
QY      499  GluGlyGlnLysIleIleTrpProSerTyrLysThrAsnGlyThr----- 513
Db      1390  ACGGAAAGATTCCCATCTCGTCGAGTTTACAGACGCTGGCAGCCGCCAAGAGTTACTAC 1449
QY      514  -----AlaThrAlaThrHis---Thr 519
Db      1450  CCGGAGGCAACAGACAGATTCACTGAGGAGTGTGATTCCTGCGCAGGTACACACA 1509
QY      520  GlnSerSerValProLeuProThrAlaThrSerMetSerSerValGlyMetAla 537
Db      1510  ACCACGATGCTCCGAAACCCGTTGAGCGGAGCTGAAGCGGATGGACCACT 1563

```

RESULT 9
 LOCUS AX534767
 DEFINITION Sequence 5 from Patent WO02068623.
 ACCESSION AX534767
 VERSION AX534767.1 GI:25261241
 KEYWORDS Aspergillus niger
 SOURCE Aspergillus niger
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE 1
 AUTHORS Edens, L., dijk Van, A.A., Krubasik, P., Albermann, K., Stock, A., Kimpel, E., Klubbauer, S., Wagner, C., Fritz, A., gustedt Von, W., Heinrich, O., Maier, D., Spreafico, F., Folkers, U., Hopper, S., Kemmer, W., Tan, P., Stiebler, J. and Albang, R.
 TITLE Novel genes encoding novel proteolytic enzymes
 JOURNAL Patent: WO 02068623-A 5 06-SEP-2002;
 DSM N.V. (NL)

FEATURES
 source
 1. 2940
 /organism="Aspergillus niger"
 /mol_type="genomic DNA"
 /db_xref="taxon:5061"

BASE COUNT 691 a 818 c 732 g 699 t

ORIGIN

Alignment Scores:
 Pred. No.: 7.64e-61 Length: 2940
 Score: 894.00 Matches: 222
 Percent Similarity: 49.84% Conservative: 81
 Best Local Similarity: 36.51% Mismatches: 201
 Query Match: 30.97% Indels: 105
 DB: 6 Gaps: 16

US-09-712-338-2_COPY_19_555 (1-537) x AX534767 (1-2940)

QY 9 ValGlyArgArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrAlaAsn 28
 Db 498 GTCTCCCGCGGCCAGTGTGGCTCCGCCACGAGTCTCATTCACCACGAGGATATCTC 557
 QY 29 AsnValThrIleArgTyrLysGluProGlyAlaGluGlyValCysGluThrProGly 48
 Db 558 GACATCCCGCTCCGCTACAAACAGATC---CCCACCGGCATTTGTAGACTGATCCAGT 614

Score:	887.50	Matches:	223
Percent Similarity:	48.15%	Conservative:	77
Best Local Similarity:	35.79%	Mismatches:	191
Query Match:	30.74%	Indels:	133
DB:	6	Gaps:	17
US-09-712-338-2_COPY_19_555 (1-537) x AX534809 (1-3080)			
QY	13	GlnLeuProLysAsnProThrGlyValLysThrLeuThrAla-----AsnAsnVal	30
DB	563	CAATTTCCTCCGACCGGAGGACATCACTGCTCAAGTCCAAAGTGCATGAGAAIGTG	522
QY	31	ThrIleArgTyrLysGluProGlyAlaGlu-----	40
DB	623	ACTATTCTTCAAGACGTGTGCGAGA-GTATCTAGAAATAGCTTTTATGCTCGATGC	681
QY	41	-----GlyValCysGluThrThrProGlyValLysSerTyrSerGly	54
DB	682	CGTGTGATTGTCAAGCTTGAATTTGCGAACTACGCGGGTGTCCGATCTTATTCGGC	741
QY	55	TyrValAspThrSerPro-----	60
DB	742	TATGTACACTTCCGCCCGGTTCCTTCGACGGGACAGGACAGTGCAGGATTATCT	801
QY	61	-----Glu-SerHisH	64
DB	802	ATCAACACGTAAACCAATCTGAAACATTTGGAGGATGAGCAATTACTGAGCCTCAACAG	861
QY	64	rPhePheTrpPhePheGluAlaArgHisAsnProGluThrAlaProIleThrTrpLe	84
DB	862	CTCTCTTTTGGTTTTCGAAGCCGCAAGATCCAGCAATGCGCCTTGSCCATCTGGCT	921
QY	84	uAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuLeuGlyProCysH	104
DB	922	CAATGGCGGTCGGGGTGGCTGCTGCTATGGGCTCCTTGAAGAAATAGTCTCTGTT	981
QY	104	sValAsnSer---ThrPheAspTyrIleAsnProHisSerTrpAsnGluValSerAs	123
DB	982	CATTGCATCAGATCCCAAGACACAGCTCCTCAATCTTGAGTTGGAACAATGAAGTCA	1041
QY	123	nLeuLeuPheLeuSerGlnProGlyValGlyPheSerTyrSer-----	138
DB	1042	TCCTCTATTCTTACACGCAACTCAAGTCGGCTTCTCATAGATGTCCTCCAAATGG	1101
QY	139	-----AspThrValAspGlySerIleAsnProValThrGly-----	151
DB	1102	CACCTTGTTCGACTCGGACCGGCGGAGAGAGATAGTTTCGGTCTATTCCTCATGA	1161
QY	151	lValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLe	171
DB	1162	TGTTCCCCAGTCCCAACTTCACTCCCATCATGTTGGTACCTTTGCAAGC-----	1215
QY	171	uIleAspThrThrAsn-----LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnG	188
DB	1216	TGCACAGACACTAATGGGACTGCATTCGGGGCTCAGCTCTATGGCATTTCCGCGCAAC	1275
QY	188	yPheLeuSerGlyLeuProSerLeu-----AspSerArgValGlnSerLysAspPh	205
DB	1276	CTGGTTTTCAGTTCACACTACAGCCAAACGATCATGCTGC-----	1321
QY	205	eSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTy	225
DB	1322	-AGTCTCTGGCTGAAAGTTACGGAGGCAATATGTTCCAGGCACTTTCGGTCTCTCCA	1380
QY	225	rGluGlnAsnGluArgIleAlaAsnGlySerVal---AsnGlyValGln---LeuAsnPh	243
DB	1381	ACAGCAGAAATGACAAATTCGAGAGGGGACTTCAGAGAGAGCGGTGCACAGTATTGCACT	1440
QY	243	eAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrPr	263
DB	1441	CGACCGCTTGGCATTTGACCGCTTGTATGATATGTTGATCCCAAGAGAGGCTTACAT	1500
QY	263	oGluPheAlaValAsnAsn-----	269

RESULT 11
D86560
LOCUS Schizosaccharomyces pombe cpy1 gene for carboxypeptidase Y,
DEFINITION complete cds.
ACCESSION D86560
VERSION D86560.1 GI:3046860
KEYWORDS cpy1; carboxypeptidase Y.
SOURCE Schizosaccharomyces pombe (fission yeast)
ORGANISM Schizosaccharomyces pombe

4308 bp DNA linear
PLN 14-APR-1998

	FEATURES	Location/Qualifiers
QY	389 pPheValTrpProAsnPhelIeGluAspLeuGluLurleu-----AlaLeuProVa	1..37000 /organism="Schizosaccharomyces pombe" /mol_type="genomic DNA" /strain="972h-" /db_xref="taxon:4896" /chromosome="I" /map="IR"
Db	3490 CTGGATCGCTAAGACTTTCGGTGACGATGTACACGCCAACTCTAGAACGTCGCCTACCGT	/clone="cosmid c19G12" complement(join(1..823,869..926,1018..1140,1183..1402, 1451..1688)) /gene="SPAC19G12.01c" /note="synonym: cut20+ " complement(join(1..823,869..926,1018..1140,1183..1402, 1451..1688)) /partial /gene="SPAC19G12.01c" /note="SPAC19G12.01c, len:487" /codon_start=1 /label=cut20+ /product="cyclosome/apc subunit cut20/apc4 " /protein_id="CAB53521.1" /db_xref="GI:5791350" /translation="MVSKSFYKPKKKHFEWINEINRVERVSGSLKRVILCPSEMLIAI LFCNSHLICCRSNQRINDVDPHDELCELCWNHDGNLIIVFKNGEKLIDSGEGL VEORPASRLAVIMITWAMQETIVNEKRNDLFDTATVMPLGLTPSSAKERIFPSK ATAOFFEPKREGNNKKVELLSILDGERGIYINNFSYSIGESDSLKALNLGVPI HSITNDLVHLICKGGINISLKITMPLLNKGDISVIDIATMRGHVLRVLEEV NMYTEFDNVFKSEASFICTDALYSKYSDTFPSLOELQFTMGIPSLKLEWIN ERVGDVLKANWERAWNYSTLIIFCCQEFVIPACERLIVILLSRGKSINGCHMKNTL LDALKVEDCLATLGQNVSFLNCFEEKRYMKHFISWLNIYAIVDFNISPESSIP QETIEHNETHVIYIRHLSFRSLKTSYFMGTGPLQIRD " complement(join(2169..2404,2446..3702,3774..4239, 4342..4767)) /gene="SPAC19G12.02c" /note="synonym: pms1 " complement(join(2169..2404,2446..3702,3774..4239, 4342..4767)) /gene="SPAC19G12.02c" /note="SPAC19G12.02c, len:794 " /codon_start=1 /product="DNA mismatch repair protein pms1 " /protein_id="CABI0113.1 " /db_xref="GI:2239195 " /db_xref="SWISS-PROT:P54280 " /translation="MSTVRPIDANTVHKICSGOVITDVAVKELVENSLSGATTIE IRPKNGINSIEFWDNGSIGDADGESICKUFTSKITDFDELEAQOTFGFGRGALS LCAGVOILTSTAQTONEAPKGVOLNLDHGSLKDKITLPQGTISVWMDLCTPLPRV KLEIKNKREPKSAISLLQAYATISINKEPMVYHOTKSGKTIQLNSTNMKDMKIN NVPGTVSSLIPWNGDIIEGTISRHVGSSTRASNEROMLFNRRLVNLPTARVIO VFYPYMAQSPPFAINLRITNGTDINVSDDKRSVLSLEDSEDSIIIEFNKSLONICES GHAICSRQSQISFYSSQIPDSGSTDGQELPQSPATESETSDSSFSYKRSCKPCRA LVBATQAPTSATVAEGSLAOSKPLPERLOKDSMRKSPLINEKVTAASSERMKKKLA LEASTYDTSMTKTIDSFFPLKOPINKPSSNPNNLLINDPSPATVPARTINILETS HNESYVSTLSSIPRTQTSVANRPSTALOKLNFPOSRDLDKTKSKTINITSLS VQDNIVSDALLKFSNKIGYVHIDISENGEDHLNLTVEKADFRLRMVVGQNRGFIV RKGFVAILDIQHASEDKENFHHLKSNLVINSODVLPRKDLAAETEVLIDHLLIL RKGFVAILDIQHASEDKENFHHLKSNLVINSODVLPRKDLAAETEVLIDHLLIL MLASKRRSMVMIGRALTTSEMNTIIVRHAELESKPWCNCPHGPRETHRLHLUKADI " complement(2405..2417) /gene="SPAC19G12.02c" /note="ttaactctgttag, splice branch and acceptor " complement(2440..2445) /gene="SPAC19G12.02c" /note="gtaagt, splice donor sequence " complement(join(3623..3702,3774..4239,4342..4758)) /gene="SPAC19G12.02c" /note="Match to PF01119 DNA_mis_repair, DNA mismatch repair protein Score 529.60 " complement(3703..3719) /gene="SPAC19G12.02c" /note="ctaaccaattatttag, splice branch and acceptor " complement(3768..3773)
QY	407 IargValSerLeulleTy-----GlyAspAlaAsPtyrilleCysasnTrpphedglyGlyGl	misc_feature
Db	3550 T-----CTTAICTATGCGGTGATGCTGACTACATTTGCAATTCATATGCGGAAGA	misc_feature
QY	426 nAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaglyTyrTh	misc_feature
Db	3601 ACCTTGGACCGACGACCTGTAGTGGCGTGGTTCACCGTAGTTTTATGAGCGCGAATWGAA	misc_feature
QY	446 rProLeuLysValasnglyValGluTyrGlygluThrArgGluTyrGlyAsnPheSerPh	misc_feature
Db	3661 GCCTTGGAGTCTTAATGGAAGAACGCTGCTGCTGAAGTCTTTCAAAACACTTGGTTA	misc_feature
QY	466 eIfhArqValTyrGluAlaglyHisGluValproTyrTyrGlnProileAlaserLeuGl	misc_feature
Db	3721 TCCTGCCCTCTCAAGAGCTGGTACATGTTCCCTTCATACCACCGCATGATTAGTA	misc_feature
QY	486 nLeuPheAsnArgThrIlePhegly 494	misc_feature
Db	3781 AATGTTGAACACCTGGATAGATGGT 3805	misc_feature
RESULT 12		
LOCUS	SPAC19G12/c	
DEFINITION	S.pombe chromosome I cosmid c19G12.	
ACCESSION	297209	
VERSION	297209.1 GI:2879763	
KEYWORDS	aldose reductase; carboxypeptidase Y; conserved hypothetical; cut20+; cyclosome/apc subunit cut20/apc4; dna mismatch repair protein; fatty acid hydroxylase; histone h2a-beta; h2az; Mitochondrial energy transfer protein; pcyl; phosphatidylinositol-4-phosphate; pms1; polysaccharide deacetylase; prenyl transferase; RNA-binding; RNP-1; splicing factor; trehalose-phosphatase; tricarboxylate transport; tRNA-Leu; ureidoglycolate hydrolase.	
SOURCE	Schizosaccharomyces pombe (fission yeast)	
ORGANISM	Schizosaccharomyces pombe	
REFERENCE	Oliver,K., Harris,D., Wood,V., Barrell,B.G. and Rajandream,M.A. Submitted (30-JUN-1997) Schizosaccharomyces pombe chromosome I sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk On Feb 13, 1998 this sequence version replaced gi:2239193.	
AUTHORS	1 (bases 1 to 37000).	
TITLE	Direct Submission	
JOURNAL	Submitted (30-JUN-1997) Schizosaccharomyces pombe chromosome I sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk On Feb 13, 1998 this sequence version replaced gi:2239193.	
COMMENT	Notes: Details of yeast sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/S.pombe/) Protein coding regions (CDS) have been predicted with the help of computer analysis using the GeneFinder program in Pombase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid c19G12 is overlapped at the 3' end by cosmid c23A1.	

```

/misc_feature
/gene="SPAC19G12.02c"
/note="gtatgt, splice donor sequence"
complement(4240..4253)
/gene="SPAC19G12.02c"
/note="ctaacggaatgtag, splice branch and acceptor"
complement(4336..4341)
/misc_feature
/gene="SPAC19G12.02c"
/note="gtaaagt, splice donor sequence"
complement(4468..4488)
/misc_feature
/gene="SPAC19G12.02c"
/note="PS00058 DNA mismatch repair proteins mutL"
5410..6372
/gene="SPAC19G12.03"
5410..6372
/gene="SPAC19G12.03"
/note="SPAC19G12.03, len:320, LOW SIMILARITY:Yersinia pseudotuberculosis., CAB46604, hypothetical 33.7 kD protein., (293 aa), fasta scores: opt: 155, E():0.003, (27.5% identity in 265 aa)"
/codon_start=1
/label="SPAC19G12.03"
/product="possible polysaccharide deacetylase"
/protein_id="CAB10114.1"
/db_xref="GI:2239196"
/db_xref="SPREMBL:O13842"
/translation="MYETRLDIGNAGKPVDTNPWNNKIAVSEFVNVYEGGERSLLY
EDGGFTFTFAGLGPNNRPVGRSTESCFEGSCGFWRILNLFKKHKVPTCNAL
GOAVEKNPVVGMAGEAGCEVGSISHRWINEGVPPETFEHKKSVQALOKASPSNS
APRSWTGVASLNRKLVCVQKDLGPDPDSYDNDYFVWADPLASKPGAEDDK
GLLIVPTLEVDNMKYAVAPGFCNSDDFTFYARDAFDVLYEEGLEGAPKMMTGLHCR
LTGRGRFGLQKLMEHITSKEGVVATREQLAQASAKHPYKA"
5608..6042
/misc_feature
/gene="SPAC19G12.03"
/note="Pfam match to entry PF01522 Polysac_deacet"
6888..7463
/gene="SPAC19G12.04"
6888..7463
/gene="SPAC19G12.04"
/note="SPAC19G12.04, len:191, SIMILARITY:Saccharomyces cerevisiae, DAL3_YEAST, ureidoglycolate hydrolase, (195 aa), fasta scores: opt: 360, E():5e-17, (36.5% identity in 200 aa)"
/codon_start=1
/label="SPAC19G12.04"
/product="putative ureidoglycolate hydrolase"
/protein_id="CAB10115.1"
/db_xref="GI:2239197"
/db_xref="SPREMBL:O13843"
/translation="MEAKKIYAQALIDEAFAPGVSVOOKDDVKMVSANGIAKKYLK
VSEIQNYEKSSASTRGVWNFFSTHPSVHPANDEHAAPQISVLRHPFTTITFPM
CRSSDQAYLIAVAPNPDGMPDWNQTAFAVKAQGVITYSAGVWHAPWYTGKTNL
AAFNVENGVAEEDDCVQSTESPIEFIKIST"
8397..9272
/gene="SPAC19G12.05"
8397..9272
/gene="SPAC19G12.05"
/note="SPAC19G12.05, len:291, SIMILARITY:Saccharomyces cerevisiae, TXTP_YEAST, tricarboxylate transport protein, (299 aa), fasta scores: opt: 738, E():0, (49.0% identity in 290 aa)"
/codon_start=1
/label="SPAC19G12.05"
/product="putative tricarboxylate transport protein"
/protein_id="CAB10116.1"
/db_xref="GI:2239198"
/db_xref="SPREMBL:O13844"
/translation="MSTVAIKTPKSNKPCPVKSIAGGVAGAIEISITYPAEFAKIR
LQLYRNESTKAKLPFGLEWYRGNGLSKAAVRFAPDSFKLSLSDHGHILT
GPRTVLGLGVAESVLVLTPTFESIKTAI IDDRKPNRPLKGFQASRLIVHENGIR
GLYRLGAATVAQOASVGRFTAYNSIKOSLQSLRPPDEKLTIVTFFLVGAGIIVY
YCTQPTDITVKSQMQLSASKEYKNSIHCAYKILITQDQLRWFWSGATPRLARLILSGGI
VTVTEKVMELIKPFP"
8397..9020
/misc_feature

```

```

/misc_feature
/gene="SPAC19G12.05"
/note="Pfam match to entry mito_carr PF00153, Mitochondrial carrier proteins"
8505..8531
/misc_feature
/gene="SPAC19G12.05"
/note="PS00215 Mitochondrial energy transfer proteins signature"
complement(10121..10516)

```

```

Alignment Scores:
Pred. No.:      Length:      37000
Score:          652.50      Matches:      170
Percent Similarity: 48.47%    Conservative: 67
Best Local Similarity: 34.76%  Mismatches:   168
Query Match:      22.60%     Indels:       85
DB:               8          Gaps:        15

US-09-712-338-2_COPY_19_555 (1-537) x SPAC19G12 (1-37000)

QY 31 ThrileArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVallys 50
Db 20080 ACTCTTCGCGTTAAAGATCCAAACCTGAATCACTGGTATTGACACT-----GTAAAG 20027

QY 51 SerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheTrpPhePheGlu 70
Db 20026 CAATACACCGGTATTAGATGTCGAAGATGACACACACTCTTTCTTCGTGTTTGTAA 19967

QY 71 AlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySer 90
Db 19966 TCTAGAAATGATCCCGAGATGATCCGTCGTGTGTGGTGTGAACGGTGGTCTGTGTC 19907

QY 91 AspSerLeuLeuGlyLeuPheGluGluGlyProCysHisValasn---SerThrPhe 109
Db 19906 TCTTCCCTTACTGTTGTTTCATGGAATAGTCTCTTCAATCAACATGAGACACTCTT 19847

QY 110 AspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGln 129
Db 19846 AAACCCGAATATACCCCTCACAGTTGGAACCTCAATGCTTCAGTATCTTTTTCGATCAA 19787

QY 130 ProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThr 149
Db 19786 CCTATCAACACGGGTTTCAGCAACGGAGAT-----GACTCGGTCTTGACACTGTACG 19733

QY 150 GlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAla 169
Db 19733 -----

QY 170 ThrLeuIleAspThrThrAsnLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 189
Db 19732 -----GCTGGTAAGGATGTTATGCG-ATTCCTTGAACCTTTT 19698

QY 190 LeuSerGlyLeuProSerLeuAspSerArgValGlnSer-LysAspPheSerLeuTrpTh 209
Db 19697 CTTTGGCAAGTTCCCTCA-----GTACGCTCAITTTGACACTTTCACATTGCTGG 19650

QY 209 rGluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPheTyrGluGlnAsnGln 229
Db 19649 TGAATCCTATGTGCGCAATTACATCCCCAGTTTGCACAGAAATATATGAGCATACCA 19590

QY 229 uArg-----IleAlaAsnGly---SerValAsnGlyValGlnLeuAsnPheAs 244
Db 19589 AGGTGCTCAACTCTTGTTCGCCAGCGGTATGAATGGAGAGCAATACATCAATTGTAA 19530

QY 244 nSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGln 264
Db 19529 GAGTGTCTTGATGGAATAGTTTACTGATCCTTTGGTCCATACTACTTTCACGAAA 19470

QY 264 uPheAlaValAsnAsnThrTyrGly-----IleIle-----IleIle 274
Db 19469 AATGGCTTCGAGAGCCCTTACGCTCTATTATGTCACAGAGAAATGTATGTCGATTAC 19410

QY 274 sAlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 294
Db -----

```


Db	19409	TGGTGCCTATGATACC-----TCG	19392
Qy	294	sGlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu	314
Db	19391	CGCTAACCTAATCAGTGGCTGTACAGACTGCTTTACT-----CCTGT	19347
Qy	314	uCysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAl	334
Db	19346	TTGCATGTGGCTCTTTGATATGCAATAACGCTATGATGGACCA--TTTACTAAGAC	19290
Qy	334	aglyArgGlyValTyrAspIleArgHisProTyrAspAspProThrProSerTyr--	353
Db	19289	TGGACTCAACATTTATGATATTCCTGAAGAATGCCGTGACCAAGACGACCTATGCTACCC	19230
Qy	354	-----TyrAsnLysPheLeuAlaLysAspSerValMetAspAlaIleGlyVa	369
Db	19229	CGAAACGGTGCATTTGAGAGTTACTTGAACCAAGAAATTTGTCAAGAAGCTTTGGAGT	19170
Qy	369	lAsnIleAsnTyrThrGlnSerAsnAsnAspValTyrTyrAlaPheGlnGlnThrGlyAs	389
Db	19169	TGAATACGATACACAGGATGCAATACTGAAGTAAACATTTGTTCTTTTCAAGGGTGA	19110
Qy	389	pPheValTrpProAsnPheIleGluAspLeuGluGluIleLeu-----AlaLeuProVa	407
Db	19109	CTGGATCGCTAAGACTTTCCTGACGATGTCACCGCAATCTTGAAGCTGGCCTACCCGT	19050
Qy	407	lArgValSerLeuIleTyr---GlyAspAlaAspTyrIleCysAsnTrpPheGlyGlyG	426
Db	19049	T-----CTTATCTATCCCGTGTGCTGACTACATTTGCAATACATGGCAATGA	18999
Qy	426	nAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGlyTyrTh	446
Db	18998	AGCTTGGACCGACGCACTTTCAGTGGCTGCTCAACGCTGACTTTTATGAGCGCAATGAA	18939
Qy	446	rProLeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPh	466
Db	18938	GCCTTGGAGTCCCTAATGGAAGGAAGCTGGTCTGTTGAAGTCTTTCAAAAACACTTGT	18879
Qy	466	eThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIleAlaSerLeuG	486
Db	18878	TCYTGGCTCTACGAAGCTGGACACATGGTTCCTTCAACCAACCGCAAGCTATTAGA	18819
Qy	486	nLeuPheAsnArgThrIlePheGly	494
Db	18818	AATGTTGAACAGCTGGATAGATGTT	18794
RESULT 13			
LOCUS	AX594884	1527 bp	DNA linear PAT 14-FEB-2003
DEFINITION	Sequence 538 from Patent EP1258494.		
ACCESSION	AX594884		
VERSION	AX594884.1	GI:28396526	
KEYWORDS	Saccharomyces cerevisiae (baker's yeast)		
SOURCE	Saccharomyces cerevisiae		
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
REFERENCE	1		
AUTHORS	Bauer, A., Gavin, A.C., Grandi, P., Krause, R., Kruse, U., Kuester, B., Marzioch, M., Schultz, J. and Superti-Furga, G.		
TITLE	Multi-protein complexes from eukaryotes		
JOURNAL	Patent: EP 1258494-A 538 20-NOV-2002;		
CELLZONE	AG (DE)		
FEATURES	Location/Qualifiers		
source	1. 1527		
	/organism="Saccharomyces cerevisiae"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:4932"		
BASE COUNT	439 a	300 c	319 g 469 t
ORIGIN			
Alignment Scores:	7.21e-41	Length:	1527
Pred. No.:			

Score:	636.00	Matches:	167
Percent Similarity:	46.29%	Conservative:	70
Best Local Similarity:	32.62%	Mismatches:	197
Query Match:	22.03%	Indels:	78
DB:	6	Gaps:	14
US-09-712-338-2_COPY_19_555 (1-537) x AX594884 (1-1527)			
Qy	1	LeuProGlySerThrProAlaSerValGlyArgGlnLeu-----ProLys	16
Db	127	CTTCCACAGATACACAGCAACCCCTTAATTTGAACGCTTTGAATCAGCATGATCCGCTG	186
Qy	17	AsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGlu	36
Db	187	TTTACAACATTTTATTTCTTCTGTGGACAGATTACAGTTTGAGACTTAGAAGTAGAT	246
Qy	37	ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal	56
Db	247	CCTTCTAACTAGGAAT-----GACACCGTAAACAAATGTCGGGTACATG	294
Qy	57	AspThrSerProGluSerHisThrPhePheTrpPhePheGluAlaArgHisAsnProGlu	76
Db	295	GACTATAAGGATTCACAACTTTTCTTACTGGTTTGTGAAGTAGAAGATCCCTGCT	354
Qy	77	ThrAlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu	96
Db	355	AACGACCAATATATCTTTGGTTAAATGGTGGACCTGCTTCTCCTGTTTACTGGTTG	414
Qy	97	PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis	116
Db	415	CTATTGAACATGAGCCCTCATCAATTTGGCCCGCATATGAACCAATCCCAATCCCTAT	474
Qy	117	SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer	136
Db	475	TCITTGAATAATAACGCTTCAATGATCTTCTTAGAAGCCACTCGGAGTCGGCTTTTCC	534
Qy	137	TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer	156
Db	535	TATGGTGTAT-----	543
Qy	157	PheAlaGlyValGlnGlyArgTyr-ProThrIleAspAlaThrLeuIleAspThrThrAsn	176
Db	544	-----GAAAAGTCTCCTCTACAAA	564
Qy	177	LeuAlaAlaGluAlaAlaTrpGluLeuGlnGlyPheLeuSerGlyLeuProSerLeu	196
Db	565	TTAGCAGGCAAGATGCGTACATTTCTCTGGAATTTTGTGAACTTTTCTCCTCATTTA	624
Qy	197	AspSerArgValGlnSerLysAspPheSerLeuThrGluSerTyrGlyGlyHisTyr	216
Db	625	-----CGCTCCACGATTTCCACATTCGAGCGCAATCCTATCGCAGGACATTTAT	672
Qy	217	GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer	235
Db	673	ATCCCTCAATTCACATGAGATCGTTGTCAAGAACCTCGAAGA-----	717
Qy	236	ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu	255
Db	718	-----ACGTTCATTTAACTTCAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG	765
Qy	256	AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys	274
Db	766	TGATTCAGCAGATTTATYGAACCAATGCGATCGCGGAAAGGGGGGTATCACCCCTGT	825
Qy	275	AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys	294
Db	826	CTCTCATCAGAAGAATGTGAGAAAAATGAGTAAGCTGCGAGGTCGT-----	870
Qy	295	GlnAspLeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeu	314
Db	871	-----TGCTAGGTGAACAAAGTTATGTTATGCTTCTTAATCAAGT	912
Qy	315	-----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla	332


```

Db 913 TTACCAATGATAGTCGCGCACTGCTTACGTGACTCTGCACCTTTTGGACCGGTACATTAAC 972
Qy 333 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 348
Db 973 ---ACAGACTCAACGCTATGACATTAAGAGGGCCCTGTGAAGATTAATAGTACTGAIGGT 1029
Qy 349 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 365
Db 1030 ATGTGTTATACAGGTCCTCGGTATGTCGACCAAGTATGAATTTCTGAAAGTTCAGAA 1089
Qy 366 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAsnAspValTyrAlaPhe 384
Db 1090 ACGTAGGGTCGCGAGTCATTAATATATCTGGCTGTGATATAGCTGTTCACCGGATTT 1149
Qy 385 GlnGlnThrGlyAspPheValTyrProAsnPhelIleGluAspLeuGluGluLeuAla 404
Db 1150 TTGTTTACGGCGGATGAAGTAACCA---TTTCAACAATATATGTGCTGAATTTATAAT 1206
Qy 405 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGly 424
Db 1207 CACAACATTCGGGTATTAATATATATCGGGTGAAGTATATATTTGTAATGGCTGGGA 1266
Qy 425 GlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaLalaGlnPheArgSerAlaGly 444
Db 1267 AACCATGCTTGGTCCATGAGTGGATGGAATGAICAATAAACGTAGTATCAGAGAAGGATG 1326
Qy 445 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyGluThrArgGluTyrGly 462
Db 1327 TTAAGACCATGGGTCAGTAAGAACAGAGGTGGACAGAGTCAAGAACATATGGC 1386
Qy 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482
Db 1387 CCTTTCACCTTTTGGAAATATACGATCCGGTCTATATGGTCCCTATGATCATCCAGG 1446
Qy 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 494
Db 1447 GCAAGTTTGGAAATGGTCAACAGTGGATTCGGGT 1482

```

```

RESULT 14
SCYBR139W 2016 bp DNA linear PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome II reading frame ORF YBR139W.
ACCESSION Z36008 Y13134
VERSION Z36008.1 GI:536435
KEYWORDS Saccharomyces cerevisiae (baker's yeast)
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
AUTHORS Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,
Barthe,C., Baur,A., Becam,A.M., Biteau,N., Boles,E., Brandt,T.,
Brendel,M., Bruckner,M., Bussereau,F., Christiansen,C.,
Contreras,R., Crouzet,M., Czapluch,C., Demolis,N., Delaveau,T.,
Daignon,F., Dondy,H., Duesterhus,S., Dubois,E., Dujon,B., El
Bakkoury,M., Entian,K.D., Feuermann,M., Fiers,W., Foho,G.M.,
Fritz,C., Gassehuber,H., Glansdorff,N., Goffeau,A., Grivell,L.A.,
de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,
Jacq,C., Jacques,M., Jauniaux,J.C., Joniaux,J.L., Kalliesoe,T.,
Kiesau,P., Kirchrath,L., Koetter,P., Korol,S., Liebi,S., Logghe,M.,
Lohan,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Mallet,L.,
Mannhaupt,G., Messenguy,F., Miosga,T., Molemans,F., Mueller,S.,
Nasr,F., Obermaier,B., Perea,J., Pierard,A., Piravandi,E.,
Pohl,F.M., Pohl,I.M., Potier,S., Profit,M., Purnelle,B., Ramezani
Rad,M., Rieger,M., Rose,M., Schaaff-Grestenschlager,I.,
Schrens,B., Schwarzlöse,C., Skala,J., Slonimski,P.P.,
Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R.,
Urrestarazu,A., van der Aart,Q.J., van Dyck,L., Vassarotti,A.,
Vetter,I., Vierendeels,F., Viissers,S., Wagner,G., de Wergifosse,P.,
Wolfe,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kleine,K.
Complete DNA sequence of yeast chromosome II
EMBO J. 13 (24), 5795-5809 (1994)

```

```

MEDLINE 95112788
PUBMED 7813418
REFERENCE 2 (bases 1 to 2016)
AUTHORS Becam,A.M., Herbert,C.J., Nasr,F., Slonimski,P.P. and Zagulski,M.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2016)
AUTHORS MIPS
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
European yeast chromosome II sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
PEAIDRES Location/Qualifiers
source 1..2016
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="II"
328..1854
/note="ORF YBR139W"
/codon_start=1
/protein_id="CAA85097.1"
/db_xref="GI:536436"
/db_xref="SWISS-PROT:P3B109"
/translation="MKYLNVFVLQLLSIKYASFGRAFSLEDDTTFANLDKQLKLP
QNTQCTLKLRNHHDDPLFTTSSVDYDILRLKIDVPSKGLDITVKWQSGMDYKD
SKHFFYFFESRNDPANDPIILWNGPGCSFTGLLFELGESSIGADMKPHNPYSW
NNASMTFLQPLGVGFSGYDEKVSSTKLADKAYIFLELFEAFPHLRNDFHAGE
SYAGHYIPQIAHEIVVKNPRTFNLTSMVNGITDPLIOADYEPMACGKGYHPVL
SSECEKMSKAAGRCRLKLCYKSKSLPCIVATYCDLLEPYINTGNYVDIRG
PCDNDTDCMCTGLRVDOYMNPEVEQETLGSVDHNYSGCDNDYFTGFLTGDSKP
FQOYIAELNNHINIPVLVYAGDKTICNMGNHWSNELEWINKRRIORMLRPWWSKE
TGEELGVKNYGPFTFLRYDAGHWVPYDQPEASLEWVNSWISGNRAFSDLSTLENAS
"
```

BASE COUNT 608 a 379 c 392 g 637 t

ORIGIN

Alignment Scores:

Pred. No.: 1,02e-40 Length: 2016
Score: 636.00 Matches: 167
Percent Similarity: 46.29% Conservative: 70
Best Local Similarity: 32.62% Mismatches: 197
Query Match: 22.03% Indels: 78
DB: 8 Gaps: 14

US-09-712-338-2_COPY_19_555 (1-537) x SCYBR139W (1-2016)

```

Qy 1 LeuProGlySerThrProAlaSerValGlyArgGlnLeu-----ProLys 16
Db 454 CTCCACAGATACACAGCAACCCCTTAAATGGACCGTTTGAATCAGCATGATCGCTG 513
Qy 17 AsnProThrGlyValThrLeuThrAlaAsnValThrIleArgTyrLysGlu 36
Db 514 TTTACAACATTTAATCTCTCTGTCGACACAGATTACAGCTTTGAGACATTAGACAGTAGAT 573
Qy 37 ProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrVal 56
Db 574 CCTTCTAAACTAGGAATT-----GACACCGCTAAACAAATGCTGGGTATACATG 621
Qy 57 AspThrSerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGlu 76
Db 622 GACTATAGAGATCCCAACACTTTTTTACGGTTTTTGAAGTAGAGACGATCCCTGCT 681
Qy 77 ThrAlaProLethrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 96
Db 682 AACGACCCCAATATTCTTGGTTAAATGSGGACCTGTTCTTCTCTGTTTACTGGGTG 741
Qy 97 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 116
Db 742 CTATTGTAAGTAGCCCTCATCAATGGCGCATATGAACCAATCCCAATCCCTAT 801
Qy 117 SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 136

```

Db 802 TCTTGAATAATACGCTTCAATGATCTTCTTAGAACAGCCACTCGGAGTCGGCTTTC 861
QY 137 TysSerAspThrValAspGlySerIleAsnProValThrGlyValGluAsnSerSer 156
Db 862 TATGGTGAT----- 870
QY 157 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrAsn 176
Db 871 -----GAAAAAGCTCCTCTACAAAA 891
QY 177 LeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeu 196
Db 892 TTACGAGCAAGATGGTACATTTCTCGAATTTGTTTGAAGCTTTTCTCTCATTTA 951
QY 197 AspSerArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyHisTyr 216
Db 952 -----CCTCCACGATTTCCACATTCGACGGCAATCTATGAGGACATAT 999
QY 217 GlyProAlaPhePheAsnHisPheTyrGluGlnAsn---GluArgIleAlaAsnGlySer 235
Db 1000 ATCCCTCAATTCACATGAGATCGTTGTCAAGAACCTGAAAGA----- 1044
QY 236 ValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGlu 255
Db 1045 -----ACGTTCAATTTAACTTCAGTTATGATGTAATGATATACACAGACCT 1092
QY 256 AlaIleGlnAlaProTyrTyrProGluPheAlaVal---AsnAsnThrTyrGlyIleLys 274
Db 1093 TTGATTCACGAGATTTATTAAGAACCAATGCGATCGGGAAAGGGGCTATCACCCCTGT 1152
QY 275 AlaValAsnGluThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCys 294
Db 1153 CTCCTCATCAGAAAGATGTGAGAAATGAGTAAAGCTGCAGGTCTG----- 1197
QY 295 GluAspLeuIleSerThrCysLysLysLysLysLysLysLysLysLysLysLysLys 314
Db 1198 -----TGTCTGAGTTGACCAAGTATCTTATGCTTCTCAATCTCAAGT 1239
QY 315 -----CysAlaGluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAla 332
Db 1240 TTACATGCTAGTCGCGACTGCTTACTGTGACTCTGCACCTTTGGACCGTACATTAAC 1299
QY 333 PheAlaGlyArgGlyValTyrAspIleArgHisProTyrAspAspPro----- 348
Db 1300 ---ACAGGACTCAACGCTATGACATTAAGAGGCGCTGTGAAGATAATAGTACTGATGCT 1356
QY 349 -----ThrProSerTyrTyrAsnLysPheLeuAlaLysAspSerValMetAsp 365
Db 1357 ATGTGTTATACAGTCTCCGCTATGTCACCATATATGAATTTTCTGAAAGTCAAGAA 1416
QY 366 AlaIleGlyValAsnIle---AsnTyrThrGlnSerAsnAspValTyrTyrAlaPhe 384
Db 1417 ACGTAGGTCGCGAGTCATATAATTTCTGGCTGTGATATGACGTGTTCCACCGATTT 1476
QY 385 GlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluGluIleLeuAla 404
Db 1477 TTGTTTACGGCGGATGAAGTAACCA---TTTCAACAATATATGCTGAATTAAT 1533
QY 405 LeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTyrPheGly 424
Db 1534 CACACATTCGGTATTAATATATATATGCGGTGATAGGATTAATTTGTAATGCTGGGA 1593
QY 425 GlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAlaGly 444
Db 1594 AACCATGCTGTGTCATAGTTGGAATGGATCAATAAAGCTAGTATACAGAAAGGATG 1653
QY 445 TyrThrPro-----LeuLysValAsnGlyValGluTyrGlyThrArgGluTyrGly 462
Db 1654 TTAAGACCATGGGTGAGTAAAGAACACAGGTGAAGAGTTGGGCAAGTCAAGAACTATGGC 1713
QY 463 AsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyrTyrGlnProIle 482
Db 1714 CCTTTCACCTTTTGGAGATATACGATCCCGTATATGCTGCTATATCATCAACCGGAG 1773

QY 483 AlaSerLeuGlnLeuPheAsnArgThrIlePheGly 494
Db 1774 GCAAGTTTGGAAATGGTCAACAGTTGGATTTCGGGT 1809
RESULT 15
AX536454 2027 bp DNA linear PAT 22-NOV-2002
LOCUS Sequence 55 from Patent WO02064766.
DEFINITION
ACCESSION AX536454
VERSION AX536454.1 GI:25262853
KEYWORDS
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1
AUTHORS Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
TITLE Bax-responsive genes for drug target identification in yeast and
JOURNAL fungus
PATENT: WO 02064766-A 55 22-AUG-2002;
JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES
source
1..2027
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
BASE COUNT 583 a 380 c 427 g 637 t
ORIGIN
Alignment Scores: 1.02e-40 Length: 2027
Pred. No.: 636.00 Matches: 167
Score: 636.00 Conservative: 70
Percent Similarity: 46.29% Mismatches: 197
Best Local Similarity: 32.62% Indels: 78
Query Match: 22.03% Gaps: 14
DB: 6
US-09-712-338-2_COPY_19_555 (1-537) x AX536454 (1-2027)
QY 1 LeuProGlySerThrProAlaSerValGlyArgArgGlnLeu-----ProLys 16
Db 627 CTCCTCACAGATACACAGAACCCCTTAATTTGGACCGTTGAATCAGATGATCGCTG 686
QY 17 AsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyrLysGlu 36
Db 687 TTTACAACTTTTATTTCTCTGTGCACACAGATTACAGTTTGAGACTTAGAACAGTAGAT 746
QY 37 ProGlyAlaGluGlyValCysLysLysThrThrProGlyValLysSerTyrSerGlyTyrVal 56
Db 747 CCTTCTAACTAGGAAT-----GACACCGTAAACAAATGGTCGGGTACATG 794
QY 57 AspThrSerProGluSerHisThrPhePheTrpPheGluAlaArgHisAsnProGlu 76
Db 795 GACTATAGGATTCACAACTTTTCTTACTGTTTGTGAAAGTAGAACAGATCCTGCT 854
QY 77 ThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeu 96
Db 855 AACGACCAATATTTCTTTGTTAAATGGTGGACCTGGTTCTCTGCTTACTGGGTG 914
QY 97 PheGluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHis 116
Db 915 CTATTTGAATAGGCGCTTCATCAATTTGGCGCGGATATGAACCAATCCACAAATCCCTAT 974
QY 117 SerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSer 136
Db 975 TCTTGAATTAATACGCTCAATGATCTTCTTAGAACACGCCACTCGGAGTCGGCTTTCC 1034
QY 137 TyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSer 156
Db 1035 TATGTGAT----- 1043
QY 157 PheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrIleIleAspThrThrAsn 176

[illegible]

Search completed: September 17, 2003, 00:48:45

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 16, 2003, 23:02:58 ; Search time 428 Seconds
(without alignments)
3086.392 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555
Perfect score: 2887
Sequence: 1 LRGSTPASVGRRLPKNP...HTQSSVPLPTATSMSSVGM 537

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1660708 seqs, 1229959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/Cgn2_1/USPTO.spool_p/US09712338/runat_16092003_144915_19194/app_query.fasta.1.711
-DB=PublishedApplications_NA -QRM=fastap -SURFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62
-TRANS-human40.cdi -LIST=45 -LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09712338.ecgn.1.1.107 -runat_16092003_144915_19194
-NCPU=6 -ICPU=3 -NO.MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Match Length	Description
1	588.5	20.4	2218 14 US-10-206-619-1 Sequence 1, Appli

2	474	16.4	1551	10	US-09-938-842A-683	Sequence 683, App
3	361	12.5	1551	13	US-10-084-018-4	Sequence 4, Appli
4	359	12.4	1697	11	US-09-796-753-39	Sequence 39, Appl
5	359	12.4	2076	10	US-09-909-320-163	Sequence 163, App
6	359	12.4	2076	10	US-09-909-088B-163	Sequence 163, App
7	359	12.4	2076	10	US-09-905-291A-163	Sequence 163, App
8	359	12.4	2076	10	US-09-902-853-163	Sequence 163, App
9	359	12.4	2076	10	US-09-907-841-163	Sequence 163, App
10	359	12.4	2076	10	US-09-907-841-163	Sequence 163, App
11	359	12.4	2076	11	US-09-904-011-163	Sequence 163, App
12	359	12.4	2076	11	US-09-906-742-163	Sequence 163, App
13	359	12.4	2076	11	US-09-906-838-163	Sequence 163, App
14	359	12.4	2076	11	US-09-907-613-163	Sequence 163, App
15	359	12.4	2076	11	US-09-907-942-163	Sequence 163, App
16	359	12.4	2076	11	US-09-904-859-163	Sequence 163, App
17	359	12.4	2076	11	US-09-909-204-163	Sequence 163, App
18	359	12.4	2076	11	US-09-904-820-163	Sequence 163, App
19	359	12.4	2076	11	US-09-904-786-163	Sequence 163, App
20	359	12.4	2076	11	US-09-906-646-163	Sequence 163, App
21	359	12.4	2076	11	US-09-906-700-163	Sequence 163, App
22	359	12.4	2076	11	US-09-903-786-163	Sequence 163, App
23	359	12.4	2076	11	US-09-902-903-163	Sequence 163, App
24	359	12.4	2076	11	US-09-903-749A-163	Sequence 163, App
25	359	12.4	2076	11	US-09-904-119-163	Sequence 163, App
26	359	12.4	2076	11	US-09-904-956-163	Sequence 163, App
27	359	12.4	2076	11	US-09-902-736-163	Sequence 163, App
28	359	12.4	2076	11	US-09-907-794-163	Sequence 163, App
29	359	12.4	2076	11	US-09-903-943-163	Sequence 163, App
30	359	12.4	2076	11	US-09-904-462-163	Sequence 163, App
31	359	12.4	2076	11	US-09-907-925-163	Sequence 163, App
32	359	12.4	2076	11	US-09-902-692-163	Sequence 163, App
33	359	12.4	2076	11	US-09-903-520-163	Sequence 163, App
34	359	12.4	2076	11	US-09-905-056-163	Sequence 163, App
35	359	12.4	2076	11	US-09-909-064-163	Sequence 163, App
36	359	12.4	2076	11	US-09-904-553-163	Sequence 163, App
37	359	12.4	2076	11	US-09-905-381-163	Sequence 163, App
38	359	12.4	2076	11	US-09-905-088-163	Sequence 163, App
39	359	12.4	2076	11	US-09-907-575-163	Sequence 163, App
40	359	12.4	2076	11	US-09-905-075-163	Sequence 163, App
41	359	12.4	2076	11	US-09-902-759-163	Sequence 163, App
42	359	12.4	2076	11	US-09-902-634-163	Sequence 163, App
43	359	12.4	2076	11	US-09-902-713-163	Sequence 163, App
44	359	12.4	2076	11	US-09-907-979-163	Sequence 163, App
45	359	12.4	2076	11	US-09-902-615-163	Sequence 163, App

ALIGNMENTS

RESULT 1
US-10-206-619-1
: Sequence 1, Application US/10206619
: Publication No. US20030108908A1
: GENERAL INFORMATION:
: APPLICANT: Korea Research Institute of Bioscience and Biotechnology
: APPLICANT: Dong Kook Pharmaceutical Co.
: APPLICANT: Rhee, Sangki
: APPLICANT: Choi, Euisung
: APPLICANT: Kang, Hyunah
: APPLICANT: Sohn, Junghoon
: APPLICANT: Bae, Junghoon
: APPLICANT: Kim, Moowoon
: APPLICANT: Agaphonov, Michael
: TITLE OF INVENTION: Hansenula polymorpha mutants and process for the preparation o
: TITLE OF INVENTION: recombinant proteins using the same
: FILE REFERENCE: 4220-116 US
: CURRENT APPLICATION NUMBER: US/10/206,619
: CURRENT FILING DATE: 2002-07-26
: PRIOR APPLICATION NUMBER: US/09/674,617
: PRIOR FILING DATE: 2001-01-03
: PRIOR APPLICATION NUMBER: PCT/KR00/00173
: PRIOR FILING DATE: 2000-03-04
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PatentIn version 3.0

```

; SEQ ID NO 1
; LENGTH: 2218
; TYPE: DNA
; ORGANISM: Hansenula polymorpha DL1
; PUBLICATION INFORMATION:
; DATABASE accession number: U67174
; DATABASE entry date: 1996-08-17
; RELEVANT RESIDUES: (1)..(2218)
US-10-206-619-1

Alignment Scores:
Pred. No.: 1,17e-59 Length: 2218
Score: 588.50 Matches: 169
Percent Similarity: 45.61% Conservatives: 70
Best Local Similarity: 32.25% Mismatches: 200
Query Match: 20.38% Indels: 86
DB: 14 Gaps: 20

US-09-712-338-2_COPY_19_555 (1-537) x US-10-206-619-1 (1-2218)
QY 1 LeuProGlySerThrProAlaSerValGlyArgGln-----Leu 14
DB 640 ATACCTCAGAACATCAAGCTCAAGTCAGCAGCCGCGAGTCACAAATACAAAGAACCTG 699
QY 15 ProLysAsnProThrGlyValLysThrLeuThrThrAlaAsnAsnValThrIleArgTyr 34
DB 700 CCGATTTTGGGACTTCAATGTTCTCAATGAGAAGTTC-TCCAACTACAAGCTGAGGGTT 758
QY 35 Lys-----GluProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyr 52
DB 759 AAGAAGACCGACCGGAGCATTTGGGACTGGACGCACACACA-----AGACAGTAC 806
QY 53 SerGlyTyrValAspThrSerProGlu---SerHisThrPhePheThrPheGluAla 71
DB 807 TCGGATACTGTGATGGAGGAGGAGACAGACATTTCTTATGGATGTTTGAGTCC 866
QY 72 ArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyProGlySerAsp 91
DB 867 AGAAATGACCGGTCAACGACCCCTGTGATTCGTGGCTCAACGGTGTCCAGGATGCTCT 926
QY 92 SerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThrPheAsp 111
DB 927 TCCTTGACTGGAATGCTTTTGAGCTCGGCTCTGCTTATCGGTCAGATCTCAAGCCCA 986
QY 112 TyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeu 131
DB 987 ATCAACACCCATATTCGTGAATTCATGCCACCTGTGATTTCTTGACCCAGCCGTC 1046
QY 132 GlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThrGlyVal 151
DB 1047 AATGTTGGATTCTCGTACTCT----- 1067
QY 152 ValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeu 171
DB 1068 -----TCCAGTCT 1076
QY 172 IleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGlyLeuSer 191
DB 1077 GTTTCCTAACAGGTGCGAGCTGGTAAGAGCTCTATCTTCTTGGAGTGTCTTCTACAG 1136
QY 192 GlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuThrThrGluSer 211
DB 1137 CAATTCCACACTTG-----CTGAAGAAGACACTTCCACATCGCCGGCGAGTCG 1184
QY 212 TyrGlyGlyHisThrGlyProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIle 231
DB 1185 TACCGTGTGCTATACATCCCACTGTTGCCCTCCAGATTCACCCATGCTGACAGA--- 1241
QY 232 AlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGly 251
DB 1242 -----TCTTTCACCTCACTTCGGTGTTCATGTTGATGTTAACGCT 1277
QY 252 IleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaValAsnThrTyr 271

```

RESULT 2

```

US-09-938-842A-683
; Sequence 683, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPL300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16

```


770 CTGACGGAAATGCTATTGGAGATGGATAATCTGATCCGAATCAATTATAGGGGGCTAT 820

263 ProGluPheAlaValAlaAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
|||||
830 GCAGAAATTC-----CTGTACCAAAATTCGCTTGTGGGTAAGAGACCAAAAAAG 877
283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
|||
878 TAC-----TTCGAGAGCAG-----TGCATGCAATGCATAGAACACATCAGG 919
303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
:::
920 AAGCAGACATCGTGTGGAGCCTTTGAAATACTGATAAACAACACTACATAGATGCGACATTACA 979
320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
980 AGTGAATCTTCCTTACTTCCAGAAATTTACAGGATGTAGTAATTTACTATACTTTTTCG--- 1036
336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
1037 -----CGGTGCGACGGAACCTCAGGATCAGCTTTTACTATGTG 1072
356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
1073 AAATTTTGTGTCACCTCCAGAGGTGAGACCAAGCCATCCACGTGGGG---AATCAGACTTTT 1129
376 SerAsnAspValTyr-----TyrAlaPheGlnThrGlyAspPheValTyrPro 393
1130 AATGATGGAAGTATAGTTGAAAGTACTTGGCAGAGATACAGTACAGTCAGTAAAGCCA 1189
394 AsnPheIleGluAspLeuGlnIleLeuAlaLeuProValArgValSerLeuIleTyr 413
1190 TGGTTAACTGAATCATGATGAATAAT-----TATAAGGTCTGATCTACAAT 1234
414 GlyAspAlaAspTyrIle-----CysAsn 421
1235 GGCCAAGTGGACATCATCGTGGCAGCTGCCCTGCAGAGCGCTCCTTGATGGCGATGGAC 1294
422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
1295 TGGAAAGGTTCCTCCAG-----GAATACAGAAGGCGAGAAAAAAGATT 1336
440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
1337 TGAAGATCTTTTAAATCTTGACAGTGAAGTGGCTGGTTAC----- 1375
453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
1376 -----ATCCGGCAAGTGGGTGACATCCATCCATAGGTAATTAATTCGAGGTGGA 1421
473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
1421 GGACATATTATACCTCATGACAGCGCTTGACAGCTTTTGACATGATGATTAATCGATTCAAT 1480
493 Phe-----GlyTyrAsp 496
1481 TATGGAAAGAGATGGGAT 1498

RESULT 4
US-09-796-753-39
; Sequence 39, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE OF INVENTION: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546


```

; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 39
; LENGTH: 1697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)...(1528)
US-09-796-753-39

```

```

Alignment Scores:
Pred. No.: 2,18e-32 Length: 1697
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 11 Gaps: 21

```

US-09-712-338-2_COPY_19_555 (1-537) x US-09-796-753-39 (1-1697)

```

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 299 CCAGGACTGAACATGAAGAGTTATGCGGCTTCCCTCACCGTGAATAGACTTACACAC 358
QY 63 HisThrPhePhePhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
::: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 359 AACCTCTTCTTCTGTTGTTCTTCCACGCTCAGATACAGCCAGAGAGTGCCTTCTC 418
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuLeuPro 102
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 419 TGGCTACAGGTGGCGCGGAGGTTCATCCATGTTTGGACTCTTTGTGAACATGGCCT 478
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 TATGTTGTCACAAAGTAACATGACCTGGGTGACAGAGACTTCCCTCGACCAACAGCTC 538
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 ICCATGCTTTACATTCACATCCAGTGGCAGACGCTTCAGTTTCTAGTATGATACCAC 598
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 GGA----- 601
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 602 -----TATCCAGTCAATGAGGACGATGATAGCAGCGGATTTATAC 640
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 641 AGTGCACATAATTCAGTTTTC-----CAGATATTTCCTGAATATAAAAT 685
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPheAsn 222
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 AATGACITTTATGTCACTGGGGACTTATGTCAGGAAATATGTCAGCCATTCACAC 745
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 746 CTCATCCATTCCTCCCAAC-----CCTGTGAGAGGAGGATCAAC 787
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaProTyrTyr 262
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 CIGAACGGAATTGCTATTTGGAGATGATTCGTATCCGAATCAATATAGGGGGCTAT 847
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleValAsnGlnGluThrValTyrAsn 282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 GCAGAAATC-----CTGTACCAAAATGGCTTGGATGAGAGCAAAAG 895
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 896 TAC-----TTCAGAGACAG-----TCCATGAATGCATAGAACATCAGG 937
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 938 AAGCAGAACTGGTTTGGAGGCTTTGAATACTGTGATAAATACTAGATGGGACTTAA 997
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 998 AGTGATCCTTCTTACTTCCAGAAATGTACAGGATAGTAATTAATACTTACTTTTG-- 1054
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1055 -----CGGTGCACCGAACTCGAGGATCAGCATCAGCTTTACTATGTG 1090
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1091 AAATTTTGTCACTCCAGAGGTGAGACAGCCATCCACGTGGGG---AATCAGACTTTT 1147
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 393
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1148 AATGATGGAACATACTAGTGAAGAGTACTTGGAGAGAGATACAGTACAGTACAGTAA 1207
QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1208 TGGTTAACTGAATCAATGAATAAT-----TATAAGGTCTCTGATCAAT 1252
QY 414 GlyAspAlaAspTyrIle----- 421
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 GGCACACTGSACATCATCTGCGAGCTGCCCTGCAGAGCGCTCCTGTATGGCATGGAC 1312

```

```

QY 422 trpPheGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
Db 1313 TGAAGAAGTCCAG-----GAATCAAGAAGCAGAAAAAAGTT 1354
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1355 TGAAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1393
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1394 -----ATCGGCAAGCGGGTCTCCATCAGCAAGTAATATTCGAGGTGGA 1438
QY 473 GlyHisGluValProTyrGlnProIleAlaSerLeuGlnPheAsnArgThrIle 492
Db 1439 GGACATATTTTACCTATGACAGCCCTCGAGAGCTTTTGACATGATTATCGATTCAAT 1498
QY 493 Phe-----GlyTrpAsp 496
Db 1499 TATGGAAGAGTGGAT 1516

RESULT 5
US-09-909-320-163
; Sequence 163, Application US/09090320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089

```

```

; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-909-320-163

Alignment Scores:
Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 10 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-909-320-163 (1-2076)
QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 295 CCAGGACTGACATGAGAGTTATCGCGCTTCCTACCGTGATATACACTTACACAGC 354
QY 63 HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
Db 355 AACCTCTCTCTCGTTCTTCCAGCTCAGATACAGCCAGAGATGCCAGTAGTTCTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluGlyPro 102
Db 415 TGGCTACAGGTGGCGCGGAGGTTTCATCCACTTTTGGACTCTTTTGGAAACATGGGCT 474
QY 103 CysHisValAsnSerThrPheAspAspIlyrIleAsnProHisSerTyrAsnGluValSer 122
Db 475 TATTTGTGCAAGTACATGACCTTGGTGACAGACATTCCTCCCTGGACCACACGCTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
Db 535 TCCATGCTTTTACATTCACATCCAGTGGGCACAGGCTTCAGTTTACTGTATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 182
Db 598 -----TATCCAGTCAATGAGGACGATGTAGCAGCGGATTTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db 637 AGTGCACATAATTCAGTTTTC-----CAGATATTTCCTGGAATATAAAAA 681
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 222
Db 682 AATGACTTTTATGACTGGGAGTCTTATCCAGGGAATATGTGCCAGCCATTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db 742 CTCATCCATTCCCTCAAC-----CCTGTGAGAGAGTGAAGATCAAC 783

```

```

QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 784 CTGAACGGAATGCTATTGGAGATGGATATCTGATCCCGAATCAATTAAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleIleValAsnGluThrValTyrAsn 282
DB   |||||||
DB 844 GCAGAAATTC-----CTGTACCAAAATGGCTGTGGTGGATGAGAACCAAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 892 TAC-----TTCAGAACGAG-----TGCATGATGATCATAGAACATCAGG 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 934 AAGCAGAACTGGTTGGAGCCTTTGAATACACTGATTAAGTACAGCCACCTTACACA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 994 AGTATCCTCTTACTTCCAGATGTTACAGGATGTAGTAATTAATACTATTG--- 1050
QY 336 ArgGlyValTyrAspIleAlaGHisProTyrAspAspProThrProProSerTyrTyrAsn 355
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1051 -----CGGTGCACGGAACCTGAGGATCAGCTTACTATG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
DB   |||||||
DB 1087 AAATTTTGTCACTCCAGAGGTGAGACAGCCATCCACGTGGG---AATCAGACHTTT 1143
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValIrrPro 393
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1144 AATGATGGAACACTAGTTGAAAGTACTTCCGAGAGATACAGTACAGTCAGTTAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1204 TGGTTRACTGAATCATGAATAT-----TATAAGGTCTGATCTACAAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1249 GGCACACTGGACATCATCGTGGCAGCTGCCCTGCACAGAGCGCTCTGTGATGGCATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1309 TGGAAAGGATCCAG-----GAATACAAAGAGGAGCAAAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1351 TGAAGATCTTTAACTGCACAGTGAAGTGGCTGGTTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValIleGluAla 472
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1390 -----ATCCGGCAAGCGGTGACTTCCATCAGTAAATATTCAGGTGGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
DB   |||||||
DB 1435 GCACATATTTCCCTATGACAGCCTCTGAGAGCTTTTGACATGATTAATCATTCATT 1494
QY 493 Phe-----GlyTrpAsp 496
DB   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
DB 1495 TATGGAAGGATGGAT 1512

```

RESULT 6

```

US-09-909-088B-163
; Sequence 163, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen

```

```

; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,088B
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-909-088B-163

Alignment Scores:
Pred. No.: 2,99e-32
Score: 359.00
Percent Similarity: 39.51%
Best Local Similarity: 26.75%
Query Match: 12.44%
DB: 10
Length: 2076
Matches: 130
Conservative: 62
Mismatches: 178
Indels: 116
Gaps: 21

```

US-09-712-338-2_COPY_19_555 (1-537) x US-09-909-088B-163 (1-2076)

QY	47	ProGly-----VallysSerThrSerGlyTyr-----ValAspThr-SerProGluSer	62
DB	295	CCAGGACTCAACATGAAGAGTTATGCCGGCTTCCTCACCGTGAATAAGACITACACAGC	354
QY	63	HISThrPhePheThrPhePheGluAlaArgHISAsnProGluThrAlaProIleThrLeu	82
DB	355	ARCCCTCTCTCTCGGTCTCTCCAGCTCAGATACAGCCAGAGAAGATGCCCCAGTAGTCTC	414
QY	83	TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro	102
DB	415	TGGCTACAGGTGGCGGAGAGTTCATCCATGTTGTGGACTCTTTGTGGAACATGGGCT	474
QY	103	CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerThrAsnGluValSer	122
DB	475	TATGTTGTCACAGTAACATGACCTGGCTGACAGAGACTTCCCTGGACCAACAGCTC	534
QY	123	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp	142
DB	535	TCCATCTTTACATTGACAATCCAGTGGGCACAGGCTTCAGTTTACTGATGATACCCAC	594
QY	143	GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly	162
DB	595	GGA-----	597
QY	163	ArgTyrProThrIleAspAlaThrIleAspThrThrAsnLeuAlaAlaGluAlaAla	182
DB	598	-----TATGCAGCTCAATGAGGACGATGTAGCAGCGGATTTATAC	636
QY	183	TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer	202
DB	637	AGTGCACATAATTCAGTTTTC-----CAGATATTCCTGAATATAAAAT	681
QY	203	LysAspPheSerLeuThrPthrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn	222
DB	682	AATGACTTTTATGTCACCTGGGAGTCTTATGCAGGGAAATATGTGCCAGCCATGTCACAC	741
QY	223	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn	242
DB	742	CTCATCCATCCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAAC	783
QY	243	PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr	262
DB	784	CTGACCGAATCTCTATTCGAGATGGATATTCGATCCCGAATCAATATTAGGGGCTAT	843
QY	263	ProGluPheAlaValAsnAsnThrTyrGlyIleIleYsAlaValAsnGluThrValTyrAsn	282
DB	844	GCAGAAATC-----CTGTACCAAAATGGCTTGTGGATGAGAAGCAAAAAAAG	891
QY	283	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	302
DB	892	TAC-----TTCAGAGACAG-----TGCCATGAATGCATGATAGACACATCAGG	933
QY	303	GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr	319
DB	934	AAGCAACACTGGTTTGAGGCTTTGAATACTGGATAAATCTAGATGGCGATTAAACA	993
QY	320	Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly	335
DB	994	AGTGATCCTCTTACTTCCAGAATGTTACAGGATGTAGTAATTAATACTAACTTTTGTG	1050
QY	336	ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn	355
DB	1051	-----CGGTGCACGGAAACCTCAGGATCAGCTTACTACTGTG	1086
QY	356	LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnIleThrGln	375
DB	1087	AAATTTTGTCACTCCAGAGGTGAGACAGCCATCCACGTGGGG---ATNCAGACTTTT	1143
QY	376	SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro	393
DB	1144	AATGATGGAACATATAGTTGAAAGAGTACTTGGAGAAGATACAGTACAGTCAGTAAAGCCA	1203

RESULT.7

```

US-09-905-291A-163
; Sequence 163, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertschen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08

```

```

; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-905-291A-163

```

Alignment Scores:

```

Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 10 Gaps: 21

```

US-09-712-338-2_COPY_19_555 (1-537) x US-09-905-291A-163 (1-2076)

```

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
DB 295 CAGGACTGAACATGAAGAGTATCCCGCTTCCTCAGCGTGAATAAGACTTACACACG 354
QY 63 HisThrPhePheTyrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
DB 355 AACCTCTCTCTCTGTTCTCCAGCTCAGATACAGCCAGAGATGCCCGAGTAGTTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
DB 415 TGGCTACAGGTGGCGCGGAGGTTCATCCATGTTTGGACTCTTCTGGAACAIGGGCT 474
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
DB 475 TATGTTGTCAACAGTAACATGATGACCTTGGCTGACAGAGACTTCCCTGGACCAACAGGCTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
DB 535 TCCATGCTTTACATGACAAATCCAGTGGCGACAGCTTCAGTTTACTGATGATATCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
DB 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182
DB 598 -----IATGCACTCAATGAGGAGCATGTAGCACCGGATTTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
DB 637 AGTGCACATTAATTCAGTTTTC-----CAGATATTTCTTGAATATAAAAT 681

```

```

QY 203 LysAspPheSerLeuTyrThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn 222
DB 682 AATGACTTTTATGTCACCTGGGAGTCTTATGACGGAAATATATGCCAGCATTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
DB 742 CTCATCCATTCCTCAAC-----CCTGGAGAGAGGTGAAGATCAAC 783
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleAspGluAlaIleGlnAlaProTyrTyr 262
DB 784 CTGAACGGAATTCATATGGAGATGGAATTCATCCCGAATCAATATATAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
DB 844 GCAGAAATC-----CTGTACCAAAITGGCTTGTGGATGAGACGAAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
DB 892 TAC-----TTCAGAGACAG-----TGCATGATGATGATGATGATGATGATGATGAT 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
DB 934 AAGCAGAACTGTTTGGAGCTTTTGAATACTGGAIAAACTACTAGATGCGACTTAAACA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
DB 994 AGTGATCCTTCTTACTTCCAGAAATGTACAGGATGTAGTAATTAATACTACTTTTTC 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspProThrProProSerTyrTyrAsn 355
DB 1051 -----CGGTGCACGGAACCTGAGGATCAGCTTACTATGTC 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
DB 1087 AATTTTGTCTACTCCAGAGGTGAGACAAAGCATCCACGTGGG---AATCAGACTTIT 1143
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGluThrGlyAspPheValTyrPro 393
DB 1144 AATGATGGAACATACTGTGAAAGTACTTCGAGAGATACAGTACAGTCAGTTAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
DB 1204 TGGTTAACTGAATCATGATATAT-----TATAAGGTTCGTGATCTCAAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
DB 1249 GSCCAACTGGACATCATCTGGCAGCTGCCCTGACAGAGCGCTCCTTGATGGCATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln----- 439
DB 1309 TGAAGAAGATCCAG-----GAATACAGAAGAGCAGAAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
DB 1351 TGAAGACATCTTAAATCTGACAGTGAAGTGGCTGGTTC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
DB 1390 -----ATCCGCAAGCGGGTGACTTCCATCAGGTAAATTTTCAGGTGGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
DB 1435 GGACATATTTTACCCTATGACAGCTCTGAGAGCTTTTGACATGATTAATGATCAATT 1494
QY 493 Phe-----GlyTyrAsp 496
DB 1495 TATGAAAAGAGTGGAT 1512

```

RESULT 8

```

US-09-902-853-163
; Sequence 163, Application US/0902853
; Publication No. US20020192659A1
; GENERAL INFORMATION:

```

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavir, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/902,853
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US/09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 163
LENGTH: 2076
TYPE: DNA
ORGANISM: Homo Sapien
US-09-902-853-163

Alignment Scores: 2.99e-32 Length: 2076
Pred. No.:

Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservatives: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 10 Gaps: 21
US-09-712-338-2_COPY_19_555 (1-537) x US-09-902-853-163 (1-2076)
QY 47 ProGly-----VallySerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
DB 295 CCAGGACTGAACATGAAGATTATCGCGGTTCTCACCCTGTAATAAGACTTACAACAGC 354
QY 63 HisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
DB 355 AACCTCTCTCTCGTCTCTCCAGCTCAGATACAGCCAGAGATGCCAGTAGTATCTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
DB 415 TGGCTACAGGTGGCGGGAGGTTCATCCATGTTGGACTCTTTGGGACATGGGCTT 474
QY 103 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer 122
DB 475 TATGTTGTCAACAAGTAACATGACCTTGCSTGCAGAGACTTCCTCCCTGGACCAACGCTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
DB 535 TCCATGCTTTACATTCACATCCAGTGGGACAGGCTTCAGTTTCTTACTGAATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGluGly 162
DB 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla 182
DB 598 -----TATGCAGTCAATGAGGACGATGATAGCAGCGGATTATATAC 636
QY 183 TrpGluLeuLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
DB 637 AGTGCACATAATTCAGTTTCT-----CAGATATTTCTCGAATATAAANAAT 681
QY 203 LysAspPheSerLeuTrpGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn 222
DB 682 AATGACATTTATGTCACCTGGGACTCTTATCCAGGGAATATGTGCCAGCATATGCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
DB 742 CTCATCCATTCCTCAAC-----CCTGTGAGAGAGGTGAGATCAAC 783
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
DB 784 CTGAACGGAATGCTATTTGGAGATGGATATCTCTCCGAATCAATATAGGGGGGTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
DB 844 GCAGAAATC-----CTGACCAATTTGCTTTGTGGATGAGAGCAAAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
DB 892 TAC-----TTCCAGAAAGCAG-----TGCCATGAATCAATAGACACATCAGG 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
DB 934 AAGCAGAACTGGTTTGGAGGCTTTGAATACTGATAAATCTAGATAGGCGGCTTACA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
DB 994 AGTGATCCTTCTTACTTCCAGAGATTACAGGATGTAGTAATTAATACTTTTGTG--- 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspProThrProProSerTyrTyrAsn 355
DB 1051 -----CGGTGCACGGAACCTCAGGATCAGCTTACTTACTATGTG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375

```

Db 1087 AAATTTTGTCACTCCAGAGGTGACAGCAAGCCATCCAGTGGGG---AATCAGACTTTT 1143
QY 376 SerAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 393
Db 1144 AATGATGAAGTATAGTTGAAAGTACTTGGGAGAGATACAGTACAGTACAGTTAAGCCA 1203
QY 394 AsnPheLeuGluAspLeuGluGluLeuAlaLeuProValArgValSerLeuIleTyr 413
Db 1204 TGGTTAACTGAATCAATGATATAT-----TATAAGGTTCGATCTACAAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
Db 1249 GGCAAACTGGCAATCATCTGGCAGCTGCCCTGACAGAGCGCTCCITGATGGGCATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439
Db 1309 TGGNAAGATCCAG-----GAATACAGAAGAGGCGAGAAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1351 TGGAAATCTTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1390 -----ATCCGCGCAAGCGGTGACTTCCATCAGGTAATTAATTCGAGGTGGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
Db 1435 GGACATATTIACCCATGACACCGCTCTGAGAGCTTTGACATGATTAATCGAATCATT 1494
QY 493 Phe-----GlyTrpAsp 496
Db 1495 TATGGAAAAGGATGGGAT 1512

```

RESULT 9

```

US-09-907-824-163
; Sequence 163, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE OF INVENTION: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,824
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414

```

Alignment Scores:

```

Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 1.0 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-824-163 (1-2076)

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 295 CCAGGACTGAACATGAAGAGTATTCGCGGCTTCCTACCGTGAATAAGACTTACACAGC 354
QY 63 HisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrIeu 82
Db 355 AACCTCTCTCTCTGGTCTTCGCCAGCICAGATACAGCCAGAGATGCCAGTAGTTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyPro 102
Db 415 TGGCTACAGGTGGCGGGAGGTTCATCCATGTTGGACTCTTTGGGAACATGGGCT 474
QY 103 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer 122
Db 475 TATGTTGTCAAGTAAACATGACCTTGGGTGACAGAGACTTCCCTGGACCAACAGC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
Db 535 TCCATGCTTACATGACATCCAGTGGGCGACAGGCTTCAGTTTTCATGATGATACCC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db 595 GGA----- 597

```

```
QY 163 ArgTyrProThrIleAspAlaThrIleuLeuIleAspThrThrAsnLeuAlaLaGluAlaAla 182
Db 598 -----TATGCAGTCAATGAGCAGATGATGACACGGGATTTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db 637 AGTGCACATTAATCAGTTTTC -----CAGATAATTCCTGAATATAAAAT 681
QY 203 LysAspPheSerLeuThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn 222
Db 682 AATGACTTTTATGCTACTGGGAGTCTTATGACGGGAAATATGTCGACCAATTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db 742 CTCATCCATCCCTCAAC -----CCTGTGAGAGAGGTGAGATCAAC 793
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
Db 784 CTGACGGAATTCCTATGAGATGATATCTGATCCCGAATCAATATATAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleIleValAsnGluThrValTyrAsn 282
Db 844 GCAGAAATC -----CTGTACCAAAATGGCTTGTGGATGAGAGCAAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
Db 892 TAC -----TTCAGAGACAG -----TGCATCAATCATAGACACATCAGG 933
QY 303 GlnThrAsn -----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
Db 934 AAGCAGAACTGGTTGAGGCTTTGAAATCTGAAATCTAGTAACTACTAGATGGCGCTTAACA 993
QY 320 Asn -----MetCysArgAspAsnValGluGly -----ProTyrTyrAlaPheAlaGly 335
Db 994 AGTAGCTCTTCTACTCCAGATGTTACAGGAGTAGTAAATCTACTATACTTTTG --- 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
Db 1051 -----CGGTGCAGCAGCACTGAGGATCAGCTTACTATGTG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
Db 1087 AAATTTTGTCTACTCCAGAGGTGAGACAAGCCATCCAGCTGGGG ---AATCAGACTTTT 1143
QY 376 SerAsnAsnAspValIyr -----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 393
Db 1144 AATGATGGAACATVAGTTGAAAAGTACTTCGGAGAGATACAGTACAGTACAGTAAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
Db 1204 TGGTTAACTGAATCAATCAATAAT -----ATAAGGTTCTGATCTACAAT 1248
QY 414 GlyAspAlaAspIyrIle -----CysAsn 421
Db 1249 GGCAACTGGACATCATCGTGGCAGCTGCCCTGACAGAGCGCTCTCTTGATGGGATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln ----- 439
Db 1309 TGGAAAGATCCCAAG -----CAATACAAGAGGCAAGAAAAAGTT 1350
QY 440 -----PheArgSer -----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1351 TGGAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC ----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1390 -----ATCCGCAAGCGGGTGACTTCCATCAGTAAATATTTCGAGGTGGA 1434
QY 473 GlyHisGluValProTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
Db 1435 GGACATATTTTACCTATGACCAAGCCCTGAGAGCTTTTGACATGATTAATCGATTCATT 1494
```

```
QY 493 Phe-----GlyTrpAsp 496
Db 1495 TATGGAAAGAGTATGGGAT 1512
RESULT 10
US-09-907-841-163
; Sequence 163, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-907-841-163
Alignment Scores: 2.99e-32 Length: 2076
Pred. No.: 359.00 Matches: 130
Score: 39.51% Conservative: 62
Percent Similarity: 26.75% Mismatches: 178
Best Local Similarity: 12.44% Indels: 116
Query Match:
```



```

DB:          10          Gaps:          21
US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-841-163 (1-2076)

QY 47 ProGly-----ValLysSerThrSerGlyTyr-----ValAspThrSerProGluSer 62
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 295 CCAGGACTGCACATGAAGAGTATGCGGCTCTCTCCACCGTGAATAAGACTTACACAGC 354
QY 63 HisThrPheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 355 AACCTCTCTCTGCTGCTCTCCAGCTCAGATACAGCCAGAGATGCCCCAGTACTTCTC 414
QY 83 TrpLeuAsnGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 415 TGGTACAGGCTGGCGGGAGGTTCATCCATGTTGGACTCTTGTGACATCGGGCT 474
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 475 TATGTTGTCAACAAGTACATGACCTTGGTGACAGAGACTTCCCTCGACCAACAAGCTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerIleSerAspThrValAsp 142
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 535 TCCATGCTTTCATGTCAATCCAGTGGCAGAGCTTCAGTTTACTGTGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerPheAlaGlyValGlnGly 162
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 182
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 598 -----TATGCAGTCAATGAGGACGATGAGCACGGCATTTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheSerGlyLeuProSerLeuAspSerArgValGluSer 202
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 637 AGTCCACTAATTCAGTTTC-----CAGATATTCCTGAATATAAAAT 681
QY 203 LysAspPheSerLeuThrThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 222
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 682 AATGACTTTTATGTCATCGGAGCTTATCGAGGAAATATGTGCCAGCCATTTGCAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 742 CTCATCCATCCCTCAAC-----CCTGTGAGAGAGGTGAAGATCAAC 783
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 784 CTGAACGAATGCTATGAGAGTGTATTCGTCCCAATCAATATAGGGGGTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 844 GCAGAAATC-----CTGTACCAATTTGGCTTGTGGATGAGAGCAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 892 TAC-----TTCCAGAGACAG-----TGCCATGATGATGACACATCAGG 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 934 AAGCAGAACTGTTTGGAGCTTTGAATACTGATATAACTACTAGTAGCGGACTTAACA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 994 AGTGATCTTCTTACTTCCAGATGTTACAGGATGTAGTATTAATTACTATACTTTTG-- 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 1051 -----CGGTGCACGGAACCTGAGGATACAGCTTTACTATGTG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 1087 AAATTTTGTACATCCAGAGGTGAGACAGCCATCCAGTGGG-----AATCAGACTTTT 1143
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTrpPro 393

```

RESULT 11

US-09-904-011-163

; Sequence 163, Application US/09904011

; Publication No. US20030003530A1

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerber, Hanspeter

; APPLICANT: Gerriksen, Mary E.

; APPLICANT: Goddard, A.

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth, J.

; APPLICANT: Kljavin, Ivar J.

; APPLICANT: Mather, Jennie P.

; APPLICANT: Pan, James

; APPLICANT: Paoni, Nicholas F.

; APPLICANT: Roy, Margaret Ann

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wood, William, I.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; FILE OF INVENTION: Acids Encoding the Same

; FILE REFERENCE: 10466-14

; CURRENT APPLICATION NUMBER: US/09/904,011

; CURRENT FILING DATE: 2001-07-11

; PRIOR APPLICATION NUMBER: 09/665,350

; PRIOR FILING DATE: 2000-09-18

; PRIOR APPLICATION NUMBER: PCT/US00/04414

; PRIOR FILING DATE: 2000-02-22

; PRIOR APPLICATION NUMBER: US 60/143,048

; PRIOR FILING DATE: 1999-07-07

; PRIOR APPLICATION NUMBER: US 60/145,698

; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: US 60/146,222
 ; PRIOR FILING DATE: 1999-07-28
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-08
 ; PRIOR APPLICATION NUMBER: PCT/US99/20944
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313
 ; PRIOR FILING DATE: 1999-11-30
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 163
 ; LENGTH: 2076
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-09-712-338-2

Alignment Scores:
 Pred. No.: 2,99e-32 Length: 2076
 Score: 359.00 Matches: 130
 Percent Similarity: 39.51% Conservative: 62
 Best Local Similarity: 26.75% Mismatches: 178
 Query Match: 12.44% Indels: 116
 DB: 11 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-904-011-163 (1-2076)

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
 DB 295 CCAGGACTGAACATCAAGAGTTATGCCGGCTTCCTCCACCGTGAATAAGACTTACACAGC 354
 QY 63 HisthrPhePheTppPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
 DB 355 AACCTCTCTCTCTGGTCTTCCAGCTCAGATACAGCCAGAGAGTGCCTCCTCCTC 414
 QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluLeuGlyPro 102
 DB 415 TGGCTACAGGTGGCCGGAGGTTCATCCATGTTGGACCTTGTGGAAACATGGGCT 474
 QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTirPAsnGluValSer 122
 DB 475 TATGTGTCAACAGTAACATGACCTTGGCTGACAGAGACTTCCCGCIGGACCAACAGCTC 534
 QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
 DB 535 TCCATGCTTTACATTGACAAATCCAGTGGCAGAGCTTCAGTTTACTGATGATACCCAC 594
 QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
 DB 595 GGA----- 597
 QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaGluAla 182
 DB 598 -----TATGCACTCAATGAGGAGGAGTGTAGCAGGGGATTATATAC 636

QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
 DB 637 AGTCCACTAATTCAGTTTTTC-----CAGATATTCTCTGAATAATAAAT 681
 QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn 222
 DB 682 AATGACTTTTATGCTACTGGGAGTCTTATCAGGAAATATGTGCCACCATTCACAC 741
 QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
 DB 742 CTATCCATTCCTCAAC-----CCTGTGAGAGAGGTCGAGATCAAC 783
 QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTrp 262
 DB 784 CTGAACGGAATTGCTATTGGAGATCGATATCTCATCCGAATCAATATATAGGGGCTAT 843
 QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
 DB 844 GCAGAAATC-----CTGTACCAATGTGGTGTGGATGAGAGCAAAAAG 891
 QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
 DB 892 TAC-----TTCCAGAAAGCAG-----TCCCATGATGATGATAGACACATCAGG 933
 QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
 DB 934 AAGCAGAACTGGTTTGGAGCCCTTGAATACTGGATAAACTACTAGATGGCAGCTTAACA 993
 QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
 DB 994 AGTGAICCTCTTACTTCCAGAAATGTTACAGAGATGTAGTAATTAATACTTTTGTG--- 1050
 QY 336 ArgGlyValTyrAspIleArgHisProTyrAspProThrProProSerTyrTyrAsn 355
 DB 1051 -----CCGTGCAGCGAACCTGAGGATCAGCTTACTATGTG 1086
 QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
 DB 1087 AAATTTTGTCTACTCCAGAGGTGAGACAAGCAATCCAGTGGG---AATCAGACTTTT 1143
 QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 393
 DB 1144 AATGATGAACCTATAGTTCAAAAGTACTTGGAGAGATACAGTACAGTACAGTAAAGCCA 1203
 QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
 DB 1204 TGGTTAACTGAATCATGATAAT-----TATAGGTTCTGATCATCAAT 1248
 QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
 DB 1249 GGCCAACTGGACATCATCGTGGCAGCGCTGACAGAGCGCTCCTTGTATGGCATGGAC 1308
 QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGln----- 439
 DB 1309 TGGAAAGGATCCAG-----GAATACAAGAGGCAAGAAAAGTT 1350
 QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
 DB 1351 TGGAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC----- 1389
 QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
 DB 1390 -----ATCCGGCAAGCGGTGACTTCCATCAGGTAATTAATTCAGGTTGGA 1434
 QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
 DB 1435 GGACATATTTTACCTATGACCAAGCTCTGAGAGCTTTTGACATGATTAATCATTCATT 1494
 QY 493 Phe-----GlyTrpAsp 496
 DB 1495 TATGGAAGAGGATGGCAT 1512

RESULT 12
US-09-906-742-163
Sequence 163, Application US/0906742
Publication No. US20030023054A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/906,742
CURRENT FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 163

```
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-906-742-163

Alignment Scores:
Pred. No.: 2,99e-32      Length: 2076
Score: 359.00           Matches: 130
Percent Similarity: 39.51%      Conservative: 62
Best Local Similarity: 26.75%    Mismatches: 178
Query Match: 12.44%            Indels: 116
DB: 11                       Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-906-742-163 (1-2076)

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 295 CCAGGACTGAACATGAAGAGTTATCGCGGCTTCCTCACCCTGCGGAGACTTACACAGC 354
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 63 HisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
   :||: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 355 AACCTCTCTCTCTGGTCTTCTCCAGCTCAGATACAGCCAGAGATGCCCCAGTAGTCTC 414
   :||: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102
   ||||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 415 TGGCTACAGGTGGCGCGGAGGTTCTCCAGTTTGGAGCTCTTTGGAGACATGGGCCCT 474
   :||: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 TATGTTCTCACAAGTAAACATGACCTTGGTGCACAGAGACTTCCCTCGACCACACAGCTC 534
   :||: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 535 TCCAIGCTTTACATTCAGTCCAGTGGCAGCAGGCTTCAGTTTACTGATGATACCCAC 594
   :||: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 595 GGA----- 597

QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 598 -----TATGCAGTCAATCAGGACGATGAGCAGGATGAGCAGGATTTATAC 636
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 637 AGTGCACATAATTCAGTTTTC-----CAGATATTTCTCTGAATATAAAAT 681
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 222
   ||||| :||: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 682 AATGACTTTTATGTCACCTGGGAGCTTATGCGAGGAAATATGCCAGCCATTCACAC 741
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
   :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 742 CTCATCCATTCCTCAAC-----CCTGTCGAGAGAGGTGAAGATCAAC 783
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleLeuAspGluAlaIleGlnAlaProTyrTyr 262
   ||||| :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 784 CTGAACGGAATTCCTATTTGAGATGAGTATTCATCCCGAATCAATATAGGGGGCTAT 843
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValIleAsn 282
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 844 GCAGAAATC-----CTGTACCAATTTGGCTGTGTGGATGAGAGCAAAAAG 891
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 892 TAC-----TTCCAGAAGCAG-----TGCCATCAATGCATAGACACATCAGG 933
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
   :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 AAGCAGAACTGGTTTGAGGCTTTGAAATACTGATAAATCTACTAGTATGCGACTTAACA 993
   :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
   :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 994 AGTGATCCCTCTACTCCAGAAATGTACAGAGTGTAGTAATTAATCTATATAACTTTTGTG--- 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspProThrProProSerTyrIleAsn 355
Db 1051 -----CGTGCACGGAACTTGAGGATCAGCTTTACTTAAGTG 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
Db 1087 AAATTTTGTCACTCCAGAGGTGAGACAAGCCATCCACGTGGSG--AAATCAGACTTTT 1143
QY 376 SerAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValTyrPro 393
Db 1144 AATGATGGAAGTATAGTTGAAAAGTACTTGCAGAAAGATACAGTACAGTCAAGTAAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
Db 1204 TGGTTAACTGAATCAATGAATAAT-----TATAGGGTCTGTACTACAT 1248
QY 414 GlyAspAlaAspTyrIle-----CysAsn 421
Db 1249 GGCCAACTGGACATCATCGTGGCGAGCTGCCCTGCACAGAGGCTCTTGATGGCATGGAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaIleGln----- 439
Db 1309 TGGAAAGGATCCAG-----GAATACAAAGAGCGCAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1351 TGGAAAGTCTTAAATCTGACAGTGAAGTGGCGGTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1390 -----ATCCGCAAGCGGGTGACTTCCATCAGGTAAATTATTCGAGGTGGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
Db 1435 GGACATATTTTACCTATGACCAAGCTCTGAGAGCTTTTGACATGATTAATCGATTTCAT 1494
QY 493 Phe-----GlyTrpAsp 496
Db 1495 TATGGAAGAGGATGGGAT 1512

RESULT 13

US-09-906-838-163
Sequence 163, Application US/09906838
Publication No. US20030027143A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kluether, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tamas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/906.838
CURRENT FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 163
LENGTH: 2076
TYPE: DNA
ORGANISM: Homo Sapien
US-09-906-838-163

Alignment Scores:

Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-906-838-163 (1-2076)

QY 47 ProGly-----ValLysSerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 295 CCAGGACTGAACATGAAGAGTTATCGCGCTCTCCACCGTGTAATAGACTTACACACAGC 354
QY 63 HisThrPhePheTrpPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
Db 355 AACCTCTTCTTCTGTGTTCTCCAGCTCATACAGCCAGAGATGCCCGAGTATTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuGluGluGlyPro 102
Db 415 TGGCTACAGGGTGGCGGGAGGTTTCATCCAGTTTGGACTCTTTGTGACATGGGCCI 474
QY 103 CysHisValAsnSerThrPheAspTyrIleAsnProHisSerTrpAsnGluValSer 122
Db 475 TATGTGTGCACAAAGTAACATGACCTTGCGTGACAGAGACTTCCCTCGGACCAACAGGCTC 534

QY	123	AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp	144
DB	535	TCCATGCTTTTACATGACAAATCCAGCTGGCGACAGGCTTCAGTTTTCAGTGGATACCCAC	594
QY	143	GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly	162
DB	595	GGA-----	597
QY	163	ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAla	182
DB	598	-----TATGCCAGTCAATGAGGACGATGTAGCACGGGATTTATAC	636
QY	183	TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValcIuSer	202
DB	637	AGTGCACCTAAATTCAGTTTTC-----CAGATATTCCTCGAATATAAAAAAT	681
QY	203	LysAspPheSerLeuTyrThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsn	222
DB	682	AATGACITTTTAIGTCATCGGGAGTCTTAUCGAGGAAATATGCCAGCCATTTGCACAC	744
QY	223	HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValcIuLeuAsn	242
DB	742	CTCATCCATTCCTCCCAAC-----CCTGTGAGAGAGGTGAAGATCAAC	783
QY	243	PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr	262
DB	784	CTGAACGGAATTCGTATTTGGAGATGGATATCTCATCCGAATCAATATATAGGGGGCTAT	843
QY	263	ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn	282
DB	844	GCAGAAITC-----CTGTACCAAAITGGCTGTGGATGAGAGCAACAAAAAG	891
QY	283	TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLys	302
DB	892	TAC-----TTCACAGAGCAG-----TGCCATGAATCCATAGACACATCAGG	933
QY	303	GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr	319
DB	934	AAGCAGAGACTGGTTTGAGGCTTTGAAATATCTGGATAAATCTACPAGATGGCACTTAACA	993
QY	320	Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly	335
DB	994	AGTGATCCTTCTPACTTCCAGAACTTACAGGATGTAGTAATTAATACTTTTGTG---1050	
QY	336	ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn	355
DB	1051	-----CGGTGCACGGAACCTCGAGGATCAGCTTACTATGTG	1086
QY	356	LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln	375
DB	1087	AAATTTTGTCTACCCAGAGGTGAGACAGCCATCCACGTGGGG---AATCAGACATTT	1143
QY	376	SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValIrrPro	393
DB	1144	AATCATGGAATATAGTTGAAAAGTACTTCGGAGAGATACGTACAGTCCAGTTAAGCCA	1203
QY	394	AsnPheIleGluAspLeuGluIleLeuAlaLeuProValArgValSerLeuIleTyr	413
DB	1204	TGTTAACTGAAATCATGAATAAT-----TATAAGTTCGTACTCTACAAT	1248
QY	414	GlyAspAlaAspTyrIle-----CysAsn	421
DB	1249	GGCCAACTGACATCATCTCGGAGCTGCGCTGACAGAGCGCTCTGTATGGCATGGAC	1308
QY	422	TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln-----	439
DB	1309	TGGAAGGATCCACAG-----GAATACAGAGGCGCAAAAAAAGTT	1350
QY	440	-----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly	452
DB	1351	TGGAAGATCTTTAAATCTGACAGTGAAGTGGCTGGTTAC-----	1389
QY	453	ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla	472

```

Db      1390  -----ATCCGCGACGGGTGACATCCATCAGGTAATATTCGAGGTGGA 1434
Qy      473  GlyHisGluValProTyrTyrGlnProIleLeuSerLeuGlnLeuPheAsnArgThrIle 492
Db      1435  GGACAIAATTTACCTCATGACACAGCCCTCTGAGAGCTTTTGACATGATTAATCGATTCAAT 1494
Qy      493  Phe-----GlyTyrAsp 496
Db      1495  TATGAAAAGGATGGAT 1512

RESULT 14
US-09-907-613-163
; Sequence 163, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095

```

D	b		892	TAC-----TTCAGAAAGCAG-----TGCCATGAATGCAIAGAACAACATCATCAGG	933
Q	y		303	GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr	319
D	b		934	AAGCAGAACCTGGTTTGAGGCCCTTGCAATAACTAGGTAACTACTAGATCGGCACCTTAACA	993
Q	y		320	Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly	335
D	b		994	AGTAGTCCTTCCTTACTWCCAGAGATTACAGGAAGTAGTAANTACTATTAACCTTTTG---	1050
Q	y		336	ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProSerTyrTyrAsn	355
D	b		1051	-----CSGTGCACGGAACCTGAGGATCAGCTTTTACTATGTG	1086
Q	y		356	LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnLleasnTyrThrGln	375
D	b		1087	AAATTTTTGTCTACTCCAGAGGTGAGAACAGCAATCCACGIGGGG---AATCAGACTTTT	1143
Q	y		376	SerAsnAsnAspValTyr-----TyrAlaPheGlnGlnThrGlyAspPheValtppro	393
D	b		1144	AATGATCGCAACATATAGTTGAAAGTACTTCGGAGAAGATACAGTACAGTCAGTTAAGCCA	1203
Q	y		394	AsnPheIleGluAspLeuGluLleLeuAlaLeuProvalargValSerleulleTyr	413
D	b		1204	TGTTIAACTCAATCATGATAAT-----TATAAGGTTCTGATCTACAAT	1248
Q	y		414	GlyAspAlaAspTyrIle-----CysAsn	421
D	b		1249	GGCCAACCTGGACATCATCTGCGGAGCTGCCCTCACAGAGCGCTTCCTGATGGCATGGAC	1308
Q	y		422	TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaLaGln-----	439
D	b		1309	TGGAARAGGATCCACG-----GAATACAGAAGGCGAGAAAAAAAAGTT	1350
Q	y		440	-----PheArgSer-----AlaGlyTyrThrProleuLysValasnGly	452
D	b		1351	TGGAAGATCTCTTAAATCTGCACAGTGAAGTGGCTGGTTAC-----	1389
Q	y		453	ValglutyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla	472
D	b		1390	-----ATCCGGCAAGCGGGTGACTTCCATCAGCTAATATTTCGAGGTGGA	1434
Q	y		473	GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheasnArgThrIle	492
D	b		1435	GGACATATTTTACCCTAATGACCACGCTCTCGAGAGCTTTTACATGATGTAATCATGATTCAT	1494
Q	y		493	Phe-----GlyTrpasp	496
D	b		1495	TATGGRAAAGGATCGGGAT	1512
RESULT 15					
US-09-907-942-163					
; Sequence 163, Application US/09907942					
; Publication No. US20030027146A1					
; GENERAL INFORMATION:					
; APPLICANT: Genentech, Inc.					
; APPLICANT: Ashkenazi, Avi					
; APPLICANT: Botstein, David					
; APPLICANT: Desnoyers, Luc					
; APPLICANT: Eaton, Dan L.					
; APPLICANT: Ferrara, Napoleone					
; APPLICANT: Filvaroff, Ellen					
; APPLICANT: Fong, Sherman					
; APPLICANT: Gao, Wei-Qiang					
; APPLICANT: Gerber, Hanspeter					
; APPLICANT: Gerritsen, Mary E.					
; APPLICANT: Goddard, A.					
; APPLICANT: Godowski, Paul J.					
; APPLICANT: Grimaldi, Christopher J.					
; APPLICANT: Gurney, Austin L.					
; APPLICANT: Hillan, Kenneth, J.					
; APPLICANT: Kijavini, Ivar J.					

```

; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 163
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-907-613-163

Alignment Scores:
Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 11 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-613-163 (1-2076)
QY 47 ProGly-----VallySerTyrSerGlyTyr-----ValAspThrSerProGluSer 62
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 295 CCAGGACTCAACATGAAGAGTATGCGCGCTTCCTCACCGTGAATAGACTTCAACAGC 354
QY 63 HisThrPhePheThrPheGluAlaArgHisAsnProGluThrAlaProLleThrLeu 82
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 355 AACCTCTTCTTGTTCTTCCTCCAGCTCAGACACAGCCAGAGATGCCCGAGTAGTCTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuLleGlyLeuPheGluGluLeuGlyPro 102
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 415 TGCTACAGGGGGCGGGAGGTTTCATCCATGTTTGGACTCTTTGTGGACATGGCCCT 474
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 475 TANGTTGTCAACAGTAACATGACCTTGCTGCGTGACAGAGACTTCCCTGGACCAACAGCTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAsp 142
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 535 TCCATCTTTATCATGACAAATCCAGTGGGCACAGGCTTCAGTTTTCAGTATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnCly 162
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuLleAspThrThrAsnLeuAlaAlaGluAlaAla 182
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 598 -----TATGCAGTCAATGAGGACGATAGCAGCGGATTATATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 637 AGTGCCTAATTCAGTTTTC-----CAGATATTCCTCGAATATAAAAT 681
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsn 222
||||| :||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 682 AATGACTTTTATGCTACCTGGGAGTCTTATGACAGGAAATATGTGCCCGCATTTGCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
||||| :||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 742 CTCATCATTCCTCAAC-----CCTGTGAGAGAGGTGAGAGATCAAC 783
QY 243 PheAsnSerLeuGlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
||||| :||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 784 CTGAACCGAATTCGATTGGAGATGGATATTCGTATCCGAATCAATTATAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleIleYsAlaValAsnGluThrValTyrAsn 282
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 844 GCAGATTTC-----CAGTACCAATATGGCTGTGTGGATGAGACAGCAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGluMetProAsnGlyCysGlnAspLeuIleSerThrCysLys 302
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

```

APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,942
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 163
LENGTH: 2076
TYPE: DNA
ORGANISM: Homo sapiens
US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-942-163 (1-2076)

Alignment Scores:
Pred. No.: 2,99e-32 Length: 2076
Score: 359.00 Matches: 130
Percent Similarity: 39.51% Conservative: 62
Best Local Similarity: 26.75% Mismatches: 178
Query Match: 12.44% Indels: 116
DB: 11 Gaps: 21

US-09-712-338-2_COPY_19_555 (1-537) x US-09-907-942-163 (1-2076)

QY 47 ProGly-----ValLysSerThrSerGlyTyr-----ValAspThrSerProGluSer 62
Db 295 CCAGGACTGAACATGAAGAGTTATGCCGCTTCCCTCCATGAATAAGACTTTACAACAGC 354
QY 63 HisThrPheThrPheGluAlaArgHisAsnProGluThrAlaProIleThrLeu 82
Db 355 AACCTCTCTCTCTGTTCTTCCAGTCAGATACAGACAGAGATGCCCGAGTATCTC 414
QY 83 TrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyPro 102

Db 415 TGGCTACAGGIGGGCCGGAGGTTATCCAGTGTGGAGTCTTTGTGGAACATGGGCT 474
QY 103 CysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSer 122
Db 475 TATGTTGTCACACAGTAACATCACCTCCGTGACAGACACTTCCCTGGACACACAGGTC 534
QY 123 AsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerThrValAsp 142
Db 535 TCCATGCTTTACATGATCCAGTGGGACAGCTTCACTTTACTGATGATACCCAC 594
QY 143 GlySerIleAsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGly 162
Db 595 GGA----- 597
QY 163 ArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAla 182
Db 598 -----TATGCAGTCAATGAGGACGATGAGCAGGGATTATAC 636
QY 183 TrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSer 202
Db 637 AGTGCACATAATTCAGTTTTC-----CAGATAITTCCTGAATAATAAAT 681
QY 203 LysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAlaPhePheAsn 222
Db 682 AATGACTTTTATGTCACCTGGGAGCTTATGCGAGGAAATATGCGCAGCAITTCACAC 741
QY 223 HisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsn 242
Db 742 CTCATCCATCCCTCAAC-----CCTGTGAGAGAGTGAAGATCAAC 783
QY 243 PheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyr 262
Db 784 CTGACGGAATTCGTATTTGAGATGGATATCTGATCCGGAATCAATATAGGGGGCTAT 843
QY 263 ProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValTyrAsn 282
Db 844 GCAGAAITC-----CTGTACCAAAATGGCTTGTGGATGAGACAGCAAAAAG 891
QY 283 TyrMetLysPheAlaAsnGlnMetProAsnGlyCysGluAspLeuIleSerThrCysLys 302
Db 892 TAC-----TTCCAGAAAGCAG-----TGCCATGAATGCATAGAACACATCAGG 933
QY 303 GlnThrAsn-----ArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThr 319
Db 934 AAGCAGAACTGGTTTGAGGCTTTGAAATACTGGATAAACTACTAGATGCGACTTAACA 993
QY 320 Asn-----MetCysArgAspAsnValGluGly-----ProTyrTyrAlaPheAlaGly 335
Db 994 AGTATCCCTTCTTACTTCCAGAAATGTACAGGATGATGTAATTAATACTAATACTTTTG 1050
QY 336 ArgGlyValTyrAspIleArgHisProTyrAspAspProThrProProSerTyrTyrAsn 355
Db 1051 -----CGGTGACGGAACCTCAGGATCAGCTTACTATGTC 1086
QY 356 LysPheLeuAlaLysAspSerValMetAspAlaIleGlyValAsnIleAsnTyrThrGln 375
Db 1087 AAAITTTTGICACTCCAGAGGTGAGACAAAGCCATCCACGTGGG---AATCAGACTTTT 1143
QY 376 SerAsnAsnAspValTyr-----TyrAlaPheGlnGluThrGlyAspPheValIrrPro 393
Db 1144 AATGATGGAACTAATAGTGAAGACTTTCGCGAAGATACAGTACAGTACAGTAAAGCCA 1203
QY 394 AsnPheIleGluAspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyr 413
Db 1204 TGGTTAAGTGAATCATGATGAAT-----TATAGGTCTTGATCTACAAT 1248
QY 414 GlyAspAlaAspTyrIle----- 421
Db 1249 GSCCAACTGCACATCAICGTGGCAGCTGCCCTGCAGAGCGCTCCTTGTATGGGCGTAC 1308
QY 422 TrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAlaAlaGln----- 439

```
Db 1309 TGGAAAGGATCCAG-----GAATACAGAGAGGCGAGAAAAAAGTT 1350
QY 440 -----PheArgSer-----AlaGlyTyrThrProLeuLysValAsnGly 452
Db 1351 TGGAAAGATCTTTAAATCTGACAGTGAAGTGGCTGTAC----- 1389
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 1390 -----ATCGGCAAGGGGTGACTTCCATCAGGTAATTATTCGAGGTGA 1434
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnPheAsnArgThrIle 492
Db 1435 GGACATATTTTACCCCTATGACCAGCCTCGAGAGCTTTTGACATGATTATCGATTTCATT 1494
QY 493 Phe-----GlyTrpAsp 496
Db 1495 TATGGAAGAGGATGGAT 1512
```

Search completed: September 17, 2003, 01:54:21
Job time : 466 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 16, 2003, 22:53:23 ; Search time 3376 Seconds
(without alignments)
3865.967 Million cell updates/sec

Title: US-09-712-338-2_COPY_19_555

Perfect score: 2887

Sequence: 1 LPGSTPASVGRRLPKNP...HTQSSVPLPTATSMSSVGM 537

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09712338/runat_16092003_144914_19142/app_query.fasta_1.711
-DB=EST -QPM=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09712338 @CGN_1.1_3549 @runat_16092003_144914_19142 -NCPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: em_estha.*
2: em_esthum.*
3: em_esthu.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estcom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_paq.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	512	17.7	588	9	AW672518	AW672518 LG1_360_P
2	409	14.2	1049	29	CNS05XWN	AL420285 T3 end of
3	405.5	14.0	648	10	BE777028	BE777028 MY-23-G-0
4	405.5	14.0	653	14	CD258740	CD258740 pLMK023aG
c 5	403.5	14.0	907	14	CD456151	CD456151 Fg03_10q0
c 6	385.5	13.4	1002	29	CNS06ZOK	AL422586 T3 end of
7	384	13.3	1054	29	CNS06NN7	AL406985 T7 end of
8	375.5	13.0	986	29	CNS06Q80	AL410350 T7 end of
9	354	12.3	1005	29	CNS06EAK	AL394866 T7 end of
10	353	12.2	3268	11	AK032197	AK032197 Mus muscu
11	352.5	12.2	963	29	CNS06ZSE	AL421896 T3 end of
12	349	12.1	2493	11	AK019861	AK019861 Mus muscu
13	345	12.0	721	14	CA747625	CA747625 GAP13D07
14	333	11.5	1581	11	AF217508	AF217508 Homo sapi
15	332.5	11.5	712	14	CD045123	CD045123 pSHB017xN
16	332.5	11.5	729	13	BQ110016	BQ110016 VD0102C06
17	327.5	11.3	685	13	BW251033	BW251033 BW251033
18	327.5	11.3	700	13	BW246209	BW246209 BW246209
19	326	11.3	1492	11	AK017087	AK017087 Mus muscu
20	324.5	11.2	876	14	CD384438	CD384438 PTM009611
21	322	11.2	652	9	AV991679	AV991679 AV991679
22	321	11.1	879	29	CNS0750D	AL429491 cLore BA0
23	320.5	11.1	675	13	BW274166	BW274166 BW274166
24	320.5	11.1	675	13	BW302983	BW302983 BW302983
25	320.5	11.1	680	13	BW208298	BW208298 BW208298
26	320.5	11.1	684	13	BW248944	BW248944 BW248944
27	320.5	11.1	685	13	BW194502	BW194502 BW194502
28	320.5	11.1	686	13	BW260398	BW260398 BW260398
29	320.5	11.1	689	13	BW046109	BW046109 BW046109
30	320.5	11.1	692	13	BW194721	BW194721 BW194721
31	320.5	11.1	699	13	BW035379	BW035379 BW035379
32	319.5	11.1	566	13	BW276795	BW276795 BW276795
33	319.5	11.1	656	13	BW276796	BW276796 BW276796
34	319.5	11.1	702	13	BW212011	BW212011 BW212011
35	319.5	11.1	762	13	BQ165650	BQ165650 EST611519
36	318	11.0	676	14	CD044351	CD044351 pSHB014XA
37	318	11.0	700	12	H1749802	H1749802 Fg02_05d0
38	316.5	11.0	682	13	BW255378	BW255378 BW255378
39	313	10.8	485	12	B1200653	B1200653 Oia061fs.r
40	311.5	10.8	638	13	BW282449	BW282449 BW282449
41	311	10.8	708	14	CD213567	CD213567 HSL_41.F1
42	309.5	10.7	656	13	BW292110	BW292110 BW292110
43	308	10.7	804	14	CB644508	CB644508 OSUNED05M
44	307.5	10.7	580	13	BW197413	BW197413 BW197413
45	307	10.6	637	14	CA352124	CA352124 623325 NC

ALIGNMENTS

RESULT 1
AW672518
LOCUS
DEFINITION
LGI_360_F03.b1_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA
sequence.
EST 19-JUL-2000
588 bp mRNA linear
ACCESSION
AW672518
VERSION
AW672518.1
KEYWORDS
EST.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 588)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
An EST database from Sorghum: light-grown seedlings
Unpublished
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 580
POLYA-No.
Location/Qualifiers
1..588
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LGI)"
/notes="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: lambda Zap; Site_1: XhoI; Site_2: EcoRI ; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 145 a 151 c 152 g 140 t
ORIGIN
Alignment Scores:
Pred. No.: 2,09e-46 Length: 588
Score: 512.00 Matches: 96
Percent Similarity: 71.15% Conservative: 15
Best Local Similarity: 61.54% Mismatches: 41
Query Match: 17.73% Indels: 4
DB: 9 Gaps: 2
US-09-712-338-2_COPY_19_555 (1-537) x AW672518 (1-588)
QY 14 LeuProLysAsnProThrGlyValLysThrLeuThrAlaAsnAsnValThrIleArg 33
DB 128 CTCGCAAGGAAGTACATGACTACAAAGACGGCTACTGCACCAACAAATGTCTACTTGA 187
QY 34 TyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyValLysSerTyrSer 53
DB 188 TACAGAACCT-----GGTATCTGTGAGACGAGCGCTGGTGATACAGACTACTCA 238
QY 54 GlyTyrValAspThrSerProGluSerHisThrPhePheThrPheGluAlaArgHis 73
DB 239 GGATACGGTGGATCTCGCACCAATGCCACCGTCTTCTTCTGTTCTTCGAGCTAGGAAC 298
QY 74 AsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeu 93
DB 299 AATCCGCATCTGATCTGCTTCACTTTGTTGTTGATGGCGGTCCCGGATCATGATTCGCTG 358
QY 94 IleGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThrPheAspTyrIle 113
DB 359 ATTGGGCTCTTTGAAGAAACAGGTCATGCATGATAGTATGATATCTCACGGCTGAGTAC 418
QY 114 AsnProHisSerTrpAsnGluValSerAsnLeuPheLeuSerGlnProLeuGlyVal 133
DB 419 AACCCTACAGCTGGAAACATCTCTCAACATGCTATATTTCCAGCCGCTCGGCACG 478
QY 134 GlyPheSerTyrSerAspThrValAspGlySerIleAsnProValThrGlyValValGlu 153
DB 479 GGATTCAGCTACCGAGAACAGCGCTCGCGCTTGAATTCAGTCCACGAGCAATTCGCT 538
QY 154 ---AsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAsp 168
DB 539 TACAACCTGAGTGAATTACCGCTGATGGCGGTACCGCTCTTGAT 586

RESULT 2
CNS06XWN 1049 bp DNA linear GSS 06-JUL-2001
LOCUS T3 end of clone AY0AA004D06 of library AY0AA from strain CBS 6340
DEFINITION of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION AL420285
VERSION AL420285.1 GI:12203469
KEYWORDS GSS.
SOURCE Kluyveromyces thermotolerans
ORGANISM Kluyveromyces thermotolerans
REFERENCE 1 (bases 1 to 1049)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Bottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaja,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1049)
AUTHORS Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 10. Kluyveromyces thermotolerans
JOURNAL FEBS Lett. 487 (1), 61-65 (2000)
MEDLINE 20584720
PUBMED 11152885
REFERENCE 3 (bases 1 to 1049)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia lactis var. lactis, Kluyveromyces thermotolerans, Kluyveromyces candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES
Location/Qualifiers
1..1049
/organism="Kluyveromyces thermotolerans"
/mol_type="genomic DNA"
/strain="CBS 6340"
/db_xref="taxon:4916"
/clone_lib="AY0AA004D06"
/note="end : T3"
misc_feature
/note="similar to Saccharomyces cerevisiae ORF YBR139W [strong similarity to carboxypeptidase]"
/evidence=not_experimental
misc_feature
/note="similar to Saccharomyces cerevisiae ORF YMR297W [PRCl ; carboxypeptidase Y, serine-type protease]"
/evidence=not_experimental
BASE COUNT 262 a 253 c 252 g 280 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1.38e-34 Length: 1049
Score: 409.00 Matches: 114
Percent Similarity: 43.90% Conservative: 37
Best Local Similarity: 33.14% Mismatches: 137

```

Query Match: 14.17% Indels: 56
DB: 29 Gaps: 10
US-09-712-338-2_COPY_19_555 (1-537) x CNS06XWN (1-1049)

QY 1 LeuProGlySerThrProAlaSerValGly-----ArgArgGlnLeuProLysAsnPro 18
DB 109 GTCCCAAAATACCGTCAGCTATCAAAATGTCGTAAGATCGAGAGTTCGGTC 168
QY 19 ThrGlyValLysThrLeuThrAlaAsnValThrIleArgTyrLysProGly 38
DB 169 ACGATAAGCTGAGCTTCATGACACCTACGATGAGGGTAAGAGAGTGCACCTTCG 228
QY 39 AlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThr 58
DB 229 AAGCTGGGTGC-----GATTCIGTCAAGCAGTACTCGGGTACTGCGACAC 276
QY 59 SerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThrAla 78
DB 277 GAGGATTCCAAACACTCTTTTACTGGCATTTGAATCCAGAAAGCACCACCAATGAC 336
QY 79 ProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu 98
DB 337 CCCGTATTCTATGGCTGAACGGTGGTCCAGGGTCTCTTCTTACGGGGTTATTTTC 396
QY 99 GluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrp 118
DB 397 GAACCTTGGCCCTCTTCGTCGGCCCTGAGCTAAGCCGTCGCAATCCGACTCCGCG 456
QY 119 AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSer 138
DB 457 AATAACAATGACTACTGTCATATCTTGAGCAGCCCTTAGCGTGGTTTCTCTATGGI 516
QY 139 AspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPheAla 158
DB 517 GAC----- 519
QY 159 GlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAla 178
DB 520 -----GAGAGTGTGCTCCACACAAAGCGACCA 546
QY 179 AlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSer 198
DB 547 GGAAGAGTGTGTCATATCTTTCAGATGTTGTTTCCAAAGAGTTCCCA----- 594
QY 199 ArgValGlnSerLysAspPheSerLeuThrThrGluSerTyrGlyHisTyrGlyPro 218
DB 595 CAGTTTAGATCTACGACTCCATATTCAGGTGAATCATACGCTGGCCATTACATCCCA 654
QY 219 AlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGly 238
DB 655 -----GAGATGCCCAATCAATCGCTGCTCATGATCATCGAC 693
QY 239 ValGlnLeuAsnPheAsnSerLeuGlyIleLeuAsnGlyIleLeuAspGluAlaIleGln 258
DB 694 AAACCTTCAATCTACCTCCATCATGATTTGGCAACGGGATAACAGATTCTTTAGTCCAG 753
QY 259 AlaProTyrThrProGluPheAlaValAsnAsnThrTyrGlyIleLysAlaValAsnGlu 278
DB 754 TATGACTACTATGAGCCCATGGGTGGCAGGGGT---GGTTACAAGCGCGT---ATA 807
QY 279 ThrValTyrAsnTyrMetLysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIle 298
DB 808 ACTGAGSAGAAATGCGCAAAATGAGAAACCAATGCTCGC---TGCAGAGCTTTGAC 864
QY 299 SerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAla 318
DB 865 AATGCAATGACTACCTCTAGTTCACG-----TTCGCTTGTATGACGT 905
QY 319 ThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyVal 338
DB 910 GCGGCATACTGT---GAGAAATATGGCCATGAGTGCCTTACACTAAACAGGCGCTGAAGCT 966

```

```

QY 339 TyrAspIleArg 342
DB 967 TATGACATACGC 978

RESULT 3
LOCUS BE777028 648 bp mRNA linear EST 20-SEP-2000
DEFINITION MY-23-G-01 PinfestansMY Phytophthora infestans cDNA, mRNA sequence.
ACCESSION BE777028
VERSION BE777028.1 GI:10230683
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
1 (bases 1 to 648)
Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
Initial assessment of gene diversity for the oomycete pathogen
Phytophthora infestans based on expressed sequences
Fungal Genet. Biol. 28 (2), 94-106 (1999)
20056376
MEDLINE 10587472
PUBMED
COMMENT Contact: Govers F
Laboratory of Phytopathology
Wageningen University
Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES
source
1..648
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="DDR7602, Al mating type"
/db_xref="taxon:4787"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DH5-alpha"
/clone_lib="PinfestansMY"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Total
RNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."
```

BASE COUNT 161 a 174 c 168 g 145 t

Alignment Scores:

Pred No.: 1..75e-34 Length: 648
Score: 405.50 Matches: 86
Percent Similarity: 50.63% Conservative: 35
Best Local Similarity: 35.98% Mismatches: 77
Query Match: 14.05% Indels: 41
DB: 10 Gaps: 6

US-09-712-338-2_COPY_19_555 (1-537) x BE777028 (1-648)

```

QY 40 GluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyr-----ValAsp 57
DB 48 GAGGATTTTGTGACTCGACC-----AAGCAGTTGAGTGGCTACTTCAAGATAACA 98
QY 58 ThrSerProGluSerHisThrPhePheThrPheGluAlaArgHisAsnProGluThr 77
DB 99 GGCTCCAAGTCGAAGAACTACTTTTACTGGTTCTTCGAATCACGGCAGTCCCTCGACC 158
QY 78 AlaProIleThrLeuThrLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPhe 97
DB 159 GATCCCGCTGATCATCTGCTTACTTGGGGTCTCGGATCGACTCCATCTTGGCTTACTG 218
QY 98 GluGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSer 117
DB 219 CAGAAACAGGCCCGCTGCTCCGTGAGCAGGACTGTCTACICAGAGAAATCGTATTCG 278

```

/dev_stage="mycelium" /clone_lib="Wageningen/U.Maryland Phytophthora infestans EST Project" /note="Vector: pCMV-SPORT6.1; Site_1: SalI; Site_2: NotI"		161 a	176 c	170 g	146 t
BASE COUNT	ORIGIN				
Alignment Scores:					
Pred. No.:	1.77e-34	Length:	653		
Score:	405.50	Matches:	86		
Percent Similarity:	50.63%	Conservative:	35		
Best Local Similarity:	35.98%	Mismatches:	77		
Query Match:	14.05%	Indels:	41		
DB:	14	Gaps:	6		
US-09-712-338-2_COPY_19_555 (1-537) x CD258740 (1-653)					
QY	40	GlUGlyValCysGluThrThrProGlyVallySerTyrSerGlyTyr	-----ValAsp	57	
DB	53	GAGGATTTTGTGACTCGACC-----AAGCAGTTGAGTGGCTACTTCAAGATAACA	103		
QY	58	ThrSerProGluSerHisThrPhePhePheGluAlaArgHisAsnProGluThr	77		
DB	104	GGCTCCAGTCGAGAACTACTTTTACTGGTCTTCGAATCAGCGGCGAGTCCCTCGACC	163		
QY	78	AlaProIleThrLeuThrPheLeuAsnGlyGlyProGlySerAspSerLeuLeuGlyLeuPhe	97		
DB	164	GATCCGCTGATCATCTGGCTTACTGGCGTCTCTGGATCAGCTCCATCTTGGCTTTACTG	223		
QY	98	GlUGluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSer	117		
DB	224	CAGGAAAACGGCCGCTGCTCCGTGAACACGACACTTGTCTACTCAAGAAATCTTCAAGGCTTGCCA	283		
QY	118	TrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyr	137		
DB	284	TGGAACGAGCGTGGCAATGTCATGTGATCGACAGCGCGTGGTGGCTTCAGCTAC	343		
QY	138	SerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPhe	157		
DB	344	GGAGAC-----	349		
QY	158	AlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeu	177		
DB	350	-----AGCGGTGAATACACACGTCAGAGAGAGGTTGGGCGGAGCAC-----	391		
QY	178	AlaAlaGluAlaAlaTrpGluLeuGlnGlyPheLeuSerGlyLeuProSerLeuAsp	197		
DB	392	-----ATGTTCCACTTCTCGCAAGAATCTTCAAGGCTTGCCA-----	430		
QY	198	SerArgValGlnSerIleAspPheSerLeuThrGluSerTyrGlyGlyHisTyrGly	217		
DB	431	---GAGTATCAAAAACCTCCCTTCTACGCTTTCGAGAGAGATTACGCGGCCACTACGTG	487		
QY	218	ProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsn	237		
DB	488	CCCGGATCGCTCACAGAATCTTTTACTGGCACACCAGCAA-----AAGCAGGGA	535		
QY	238	GlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIle	257		
DB	536	CCTGTGGAATTAATTGGAAGGCTTTGGAATTTGGAGACGGCICACCGCACCTGAAGTG	595		
QY	258	GlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleAlaVal	276		
DB	596	CAGTACAGTACTACCGGACATGGCTACAAACAATACATACGGGCTCAAGGCTGTG	652		
RESULT 5					
LOCUS	CD456151/c	907 bp	mRNA	linear	EST 03-JUN-2003
DEFINITION	Fg03_10g07_R Fg03_AAPC_ECORC_Fusarium_graminearum_mycelium_trichothecene_product ion Gibberella zeae cDNA clone Fg03_10g07, mRNA sequence.				
ACCESSION	CD456151				
VERSION	CD456151.1 GI:31370891				

QY	118	TrpAsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyr	137		
DB	279	TGAACAGCGCTGCCATGTCATGTCGATCAGCACCGCGTGGTGGCTTCAGCTAC	338		
QY	138	SerAspThrValAspGlySerIleAsnProValThrGlyValValGluAsnSerSerPhe	157		
DB	339	GGAGAC-----	344		
QY	158	AlaGlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeu	177		
DB	345	-----AGCGGTGAATACACACGTCAGAGAGAGGTTGGCGGCGAGCAC-----	386		
QY	178	AlaAlaGluAlaAlaTrpGluLeuGlnGlyPheLeuSerGlyLeuProSerLeuAsp	197		
DB	387	-----ATGTTCCACTTCTCGCAAGAATCTTCAAGGCTTGCCA-----	425		
QY	198	SerArgValGlnSerIleAspPheSerLeuThrGluSerTyrGlyGlyHisTyrGly	217		
DB	426	---GAGTATCAAAAACCTCCCTTCTACGCTTTCGAGAGAGATTACGCGGCCACTACG	482		
QY	218	ProAlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsn	237		
DB	483	CCCGGATCGCTCACAGAATCTTTACTGGCAACACAGCAA-----AAGGAGGGA	530		
QY	238	GlyValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIle	257		
DB	531	CCTGTGGAATTAATTGGAAGGCTTTGGAATTTGGGAACGCGCTCACCGACCTGAAGTG	590		
QY	258	GlnAlaProTyrTyrProGluPheAlaValAsnAsnThrTyrGlyIleIleAlaVal	276		
DB	591	CAGTACAGTACTACCGGACATGGCTACAAACAATACATACGGGCTCAAGGCTGTG	647		
RESULT 4					
LOCUS	CD258740	653 bp	mRNA	linear	EST 23-MAY-2003
DEFINITION	pIMY023ag01r_213782 Wageningen/U.Maryland Phytophthora infestans EST Project Phytophthora infestans cDNA clone IMY023g01 5, mRNA sequence.				
ACCESSION	CD258740.1 GI:31046567				
VERSION	CD258740				
KEYWORDS	Phytophthora infestans (potato late blight agent)				
SOURCE	Phytophthora infestans				
ORGANISM	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora				
REFERENCE	1 (bases 1 to 653) Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F. Initial assessment of gene diversity for the oomycete pathogen Phytophthora infestans based on expressed sequences Fungal Genet. Biol. 28 (2), 94-106 (1999)				
JOURNAL	20056376				
MEDLINE	10587472				
PUBMED	Contact: Tyler B				
COMMENT	Tyler lab VBI 1880 Pratt Dr., Blacksburg, VA 24061, USA Tel.: 540-231-7318 Email: bmtyle@vt.edu PCR Primers FORWARD: M13 reverse 17mer at 5' end BACKWARD: M13 reverse 17mer at 5' end Plate: 023 row: G column: 01 Seq primer: M13 reverse 17mer at 5' end High quality sequence stop: 653. Location/Qualifiers 1..653 /organism="Phytophthora infestans" /mol_type="mRNA" /db_xref="taxon:4787" /clone="IMY023g01" /tissue_type="mycelium" /cell_line="DDR7602"				
FEATURES					
source					

[illegible]

seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
 1..1002
 /organism="Kluyveromyces marxianus"
 /mol_type="genomic DNA"
 /strain="CBS 712"
 /variety="marxianus"
 /db_xref="taxon:4911"
 /clone.lib="AZ0AA003B09"
 /clone.lib="AZ0AA"
 /note="end : T3"
 complement(<8..>994)
 /note="similar to Saccharomyces cerevisiae ORF YMR297w [PRC1 ; carboxypeptidase Y, serine-type protease]"
 /evidence=not experimental
 misc_feature
 complement(<11..>778)
 /note="similar to Saccharomyces cerevisiae ORF YBR139w [strong similarity to carboxypeptidase]"
 /evidence=not experimental
 BASE COUNT 267 a 259 c 240 g 235 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.38e-32 Length: 1002
 Score: 385.50 Matches: 102
 Percent Similarity: 42.59% Conservative: 36
 Best Local Similarity: 31.48% Mismatches: 123
 Query Match: 13.35% Indels: 63
 DB: 29 Gaps: 11
 US-09-712-338-2_COPY_19_555 (1-537) x CNS06ZOK (1-1002)
 QY 18 ProThrglyValLysThrLeuThrAlaAsnValThrlle----- 32
 DB 826 CCAGAGAGCTTTGGTCTACGGCTAGCGTCAACAGCAGCAATGTCGGGAAGTACCAGTIG 767
 QY 33 -----ArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVal 49
 DB 766 CGTGGAAGAAGATCAAGGACCCCTAAGGTGCTTGGCGTG-----GACCCCGACGTG 716
 QY 50 LysSerTyrSerGlyTyrValAspThrSerProGlu---SerHisThrPhePheTrpPhe 68
 DB 715 AAGCAATACCTCGGTACTGGAGCTTGAACAGCAGGACAGCACTTCTTCTACTGGTTC 656
 QY 69 PheGluAlaArgHisAsnProGluThrAlaProThrLeuThrLeuAsnGlyGlyPro 88
 DB 655 TTTGAGTCCAGAACAGACCCCAAGACACCTGTTATCTTGTGTTCAACGGTGGGCCA 596
 QY 89 GlySerAspSerLeuLeuGlyLeuPheGluGluLeuGlyProCysHisValAsnSerThr 108
 DB 595 GGGGTTCCTCTTTGACTGGTTTGTCTTCGAAATGGGTTCATTCCTTGGCGAAGAG 536
 QY 109 PheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuPheLeuSer 128
 DB 535 GTGAGCCAAATTACACCCACATCTTGGACACCAACGCTTCGGTTATCTTCTTGGAC 476
 QY 129 GlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnProVal 148
 DB 475 CAGCCAGTCAAGCTGGTTACTTCTTACTCTTCAATCA---GAAGGTGTCCTCAACACTGTT 419
 QY 149 ThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIleAsp 168
 DB 418 GCT----- 416

QY 169 AlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIleLeuGlnGly 188
 DB 415 -----GCAGTAAGGACGCTGTATGCGTCTTCTTCAATG 383
 QY 189 PheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeuTrp 208
 DB 382 TTCTTCCAGCAATTCCTCCGAGTAGCCTCT-----GCCCAAGATTTCCACATTGCT 332
 QY 209 ThrGluSerTyrGlyGlyHisTyrGlyProAlaPheAsnHisPhe-----TyrGlu 226
 DB 331 GGTGAATCTTATCGGGCCACTACATCCCTGTGTTGCTACAGAGATTTTGCACCCA 272
 QY 227 GlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeu 246
 DB 271 CAAGAGGAAGCT-----TCTTTCAACTTGACCTCTGTT 239
 QY 247 GlyIleLeuAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAla 266
 DB 238 TIGATCGIATACGGTTTGACCGCCCTTGACCGAGTACCTGTAGTAGAGCCATATGGCT 179
 QY 267 ValAsnAsnThrTyrGlyIleLysAlaValAsnGluThrValIyrAsnTyrMetLysPhe 286
 DB 178 TGTGTCAGGTGTTGACCCATCTGTGTGGACGAGGAGCAATGTGAAGCATG----- 125
 QY 287 AlaAsnGlnMetProAsnGlyCysGlnAspLeuLeuSerThrCysLysGlnThrAsnArg 306
 DB 124 TTGGAAGCGTACCAGCA---TGCTAACTGATTCATCTTGTACGAA----- 77
 QY 307 ThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAspAsnVal 326
 DB 76 -----TACGAATCTGTTTGTGTCATGTGTGCGAGCCTCCCTACTCTGTAACAGCTCAA 23
 QY 327 GluGlyProTyr 330
 DB 22 ATGGTCCATAC 11
 RESULT 7
 CNS06NN7
 LOCUS
 DEFINITION
 T7 end of clone AU0AA015E08 of library AU0AA from strain CBS 3082
 of Saccharomyces kluyveri, genomic survey sequence.
 ACCESSION
 AL406985
 VERSION
 AL406985.1
 GI:12171743
 KEYWORDS
 GSS.
 SOURCE
 Saccharomyces kluyveri
 ORGANISM
 Saccharomyces kluyveri
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE
 1 (bases 1 to 1054)
 Souciet,J.L., Algle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 2 (bases 1 to 1054)
 Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
 Gaillardin,C. and Casaregola,S.
 Genomic exploration of the hemiascomycetous yeasts: 9.
 Saccharomyces kluyveri
 FEBS Lett. 487 (1), 56-60 (2000)
 20584719
 11152884
 3 (bases 1 to 1054)
 Genoscope.
 Direct Submission
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 JOURNAL

Db	550	AAGGATGATATAGCTTTCTTCAGCTGTTTTTCAGACAGTTC	CCAGAGTACGCTTCT	---	606
QY	200	valGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyProAla	219		
Db	507	-----GGCCAGGACTTCATCGCTCGTGAGTCTTATGCTGGACACACACATTCCTG	CTG	660	
QY	220	PhePheAsnHisPhe-----TyrIuGlnAsnGluArgIleAlaAsnGlySerValAsn	237		
Db	661	TTTGCTTCTGAATCTTGAGCCACCCACCACTTCGAAAGN	-----	699	
QY	238	GlyValGlnLeuAsnPheAsnSerLeuGlyIleIle-----AsnGlyIleIleAspGlu	255		

D6		: :	
D7		: :	
D8		: :	
D9		: :	
D10		: :	
D11		: :	
D12		: :	
D13		: :	
D14		: :	
D15		: :	
D16		: :	
D17		: :	
D18		: :	
D19		: :	
D20		: :	
D21		: :	
D22		: :	
D23		: :	
D24		: :	
D25		: :	
D26		: :	
D27		: :	
D28		: :	
D29		: :	
D30		: :	
D31		: :	
D32		: :	
D33		: :	
D34		: :	
D35		: :	
D36		: :	
D37		: :	
D38		: :	
D39		: :	
D40		: :	
D41		: :	
D42		: :	
D43		: :	
D44		: :	
D45		: :	
D46		: :	
D47		: :	
D48		: :	
D49		: :	
D50		: :	
D51		: :	
D52		: :	
D53		: :	
D54		: :	
D55		: :	
D56		: :	
D57		: :	
D58		: :	
D59		: :	
D60		: :	
D61		: :	
D62		: :	
D63		: :	
D64		: :	
D65		: :	
D66		: :	
D67		: :	
D68		: :	
D69		: :	
D70		: :	
D71		: :	
D72		: :	
D73		: :	
D74		: :	
D75		: :	
D76		: :	
D77		: :	
D78		: :	
D79		: :	
D80		: :	
D81		: :	
D82		: :	
D83		: :	
D84		: :	
D85		: :	
D86		: :	
D87		: :	
D88		: :	
D89		: :	
D90		: :	
D91		: :	
D92		: :	
D93		: :	
D94		: :	
D95		: :	
D96		: :	
D97		: :	
D98		: :	
D99		: :	
D100		: :	
D101		: :	
D102		: :	
D103		: :	
D104		: :	
D105		: :	
D106		: :	
D107		: :	
D108		: :	
D109		: :	
D110		: :	
D111		: :	
D112		: :	
D113		: :	
D114		: :	
D115		: :	
D116		: :	
D117		: :	
D118		: :	
D119		: :	
D120		: :	
D121		: :	
D122		: :	
D123		: :	
D124		: :	
D125		: :</	

Qy 256 AlaIleGlnAlaPro^{Tyr}TyrProGluPheAla-ValAspAsnThrTyrGlvIleValAl 275

Db 793 GGAGAAC-----CTGCGGTITTTGGGAGCCGGAGCAATGTGACGGTAT 834

QY 306 qThrAlaLeuAlaAspTyrAlaLeuCys-----AlaGluAlaThrAsnMet-CysA 323

Db 878 -----TACGACTATNGAATCTGTTGGTCGTGTGCGCTGCTGCAATTACTGTA 927

QY	323	rAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArgClyValTyrAspIleArgH	343
		:::	
Db	928	ACAATGCCAGATGGGACCTTACCACGC---ACGGGTAAACAGCTTTATGACATTCTGA	984
QY	343	iSProTyrAspAspPro-----ThrProProSerTyrTyrAsnLysPheLeuAlaL	360

Db 985 AGGAATGTGAGGGTGAACCTGTGCTACTCAGAAATGCAATACTTGAACGAGTATYTAAT 1044

QY 360 YsAsp 361

Db 1045 TGGAC 1049

RESULT 8

CNS06Q80	CNS06Q80	986 bp	DNA	linear	GSS 05-JUL-2003
LOCUS					
DEFINITION	T7 end of class protein-coding gene				

DEFINITION	17 end of clone AW04A003H06 of library AW04A from strain CLIB 89
ACCESSION	AF410350
Yarrowia lipolytica, genomic survey sequence.	

REVISION AL410350.1 GT:12178581
KEYWORDS GSS.

SOURCE	ORGANISM
Yarrowia lipolytica	Yarrowia lipolytica
Yarrowia lipolytica	Yarrowia lipolytica

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Dipodascaceae; Yarrowia.

REFERENCE

1 (bases 1 to 986)

AUTHORS

Souciét,J.L., Aigle,M., Artigenave,F., Blandin,G.,
Baudry,A., Boudier,C., Chazotte,M., Dreyer,P.,
Favre,R., Gauthier,D., Guéhenneuc,H., Lecomte,
M., Mouton,R., Pignatelli,B., Rostaing,L., Sarrailh,
J., Schmitt,V., Tardieu,A., Thibaut,M., Viret,J.

Borotkin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingie, A., Llorente, B., Malbert, A.

Marperly, A., Neugebuse, C., Ozler-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenhach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711

PUBMED	11152876
REFERENCE	2 (bases 1 to 986)
SEQUENCE	1
SEQUENCE	2
SEQUENCE	3
SEQUENCE	4
SEQUENCE	5
SEQUENCE	6
SEQUENCE	7
SEQUENCE	8
SEQUENCE	9
SEQUENCE	10
SEQUENCE	11
SEQUENCE	12
SEQUENCE	13
SEQUENCE	14
SEQUENCE	15
SEQUENCE	16
SEQUENCE	17
SEQUENCE	18
SEQUENCE	19
SEQUENCE	20
SEQUENCE	21
SEQUENCE	22
SEQUENCE	23
SEQUENCE	24
SEQUENCE	25
SEQUENCE	26
SEQUENCE	27
SEQUENCE	28
SEQUENCE	29
SEQUENCE	30
SEQUENCE	31
SEQUENCE	32
SEQUENCE	33
SEQUENCE	34
SEQUENCE	35
SEQUENCE	36
SEQUENCE	37
SEQUENCE	38
SEQUENCE	39
SEQUENCE	40
SEQUENCE	41
SEQUENCE	42
SEQUENCE	43
SEQUENCE	44
SEQUENCE	45
SEQUENCE	46
SEQUENCE	47
SEQUENCE	48
SEQUENCE	49
SEQUENCE	50
SEQUENCE	51
SEQUENCE	52
SEQUENCE	53
SEQUENCE	54
SEQUENCE	55
SEQUENCE	56
SEQUENCE	57
SEQUENCE	58
SEQUENCE	59
SEQUENCE	60
SEQUENCE	61
SEQUENCE	62
SEQUENCE	63
SEQUENCE	64
SEQUENCE	65
SEQUENCE	66
SEQUENCE	67
SEQUENCE	68
SEQUENCE	69
SEQUENCE	70
SEQUENCE	71
SEQUENCE	72
SEQUENCE	73
SEQUENCE	74
SEQUENCE	75
SEQUENCE	76
SEQUENCE	77
SEQUENCE	78
SEQUENCE	79
SEQUENCE	80
SEQUENCE	81
SEQUENCE	82
SEQUENCE	83
SEQUENCE	84
SEQUENCE	85
SEQUENCE	86
SEQUENCE	87
SEQUENCE	88
SEQUENCE	89
SEQUENCE	90
SEQUENCE	91
SEQUENCE	92
SEQUENCE	93
SEQUENCE	94
SEQUENCE	95
SEQUENCE	96
SEQUENCE	97
SEQUENCE	98
SEQUENCE	99
SEQUENCE	100
SEQUENCE	101
SEQUENCE	102
SEQUENCE	103
SEQUENCE	104
SEQUENCE	105
SEQUENCE	106
SEQUENCE	107
SEQUENCE	108
SEQUENCE	109
SEQUENCE	110
SEQUENCE	111
SEQUENCE	112
SEQUENCE	113
SEQUENCE	114
SEQUENCE	115
SEQUENCE	116
SEQUENCE	117
SEQUENCE	118
SEQUENCE	119
SEQUENCE	120
SEQUENCE	121
SEQUENCE	122
SEQUENCE	123
SEQUENCE	124
SEQUENCE	125
SEQUENCE	126
SEQUENCE	127
SEQUENCE	128
SEQUENCE	129
SEQUENCE	130
SEQUENCE	131
SEQUENCE	132
SEQUENCE	133
SEQUENCE	134
SEQUENCE	135
SEQUENCE	136
SEQUENCE	137
SEQUENCE	138
SEQUENCE	139
SEQUENCE	140
SEQUENCE	141
SEQUENCE	142
SEQUENCE	143
SEQUENCE	144
SEQUENCE	145
SEQUENCE	146
SEQUENCE	147
SEQUENCE	148
SEQUENCE	149
SEQUENCE	150
SEQUENCE	151
SEQUENCE	152
SEQUENCE	153
SEQUENCE	154
SEQUENCE	155
SEQUENCE	156
SEQUENCE	157
SEQUENCE	158
SEQUENCE	159
SEQUENCE	160
SEQUENCE	161
SEQUENCE	162
SEQUENCE	163
SEQUENCE	164
SEQUENCE	165
SEQUENCE	166
SEQUENCE	167
SEQUENCE	168
SEQUENCE	169
SEQUENCE	170
SEQUENCE	171
SEQUENCE	172

AUTHORS
Casaregola, S., Neveglise, C., Lepingle, A., Bon, E., Feyrerol, C.,
Artiguenave, F., Wincker, P., and Gaillardin, C.

JOURNAL OF LIPOLYTIC GENOMIC EXPLORATION OF THE HEMIASCOMYCETOUS YEASTS: 17. Yarrowia

BOOKS RECEIVED
J. L. LEE, 487 (1), 93-100 (2000)
MEDLINE 20584727

11152876
2 (bases 1 to 1005)
de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Souciet, J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEBS Lett. 487 (1), 52-55 (2000)
20584718
11152883
3 (bases 1 to 1005)
Genoscope.
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1. .1005
/organism="Zygosaccharomyces rouxii"
/mol_type="genomic DNA"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone="AR0A020A09"
/clone_lib="AR0AA"
/note="end : 17"
<3. >985
/note="similar to Saccharomyces cerevisiae ORF YBR139w [
strong similarity to carboxypeptidase]
1 putative frameshift(s)"
/evidence=not_experimental
<3. >982
/note="Similar to Saccharomyces cerevisiae ORF YMR297w [
PC1 ; carboxypeptidase Y, serine-type protease]
1 putative frameshift(s)"
/evidence=not_experimental
BASE COUNT 284 a 187 c 231 g 302 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1.75e-28 Length: 1005
Score: 354.00 Matches: 117
Percent Similarity: 41.62% Conservative: 42
Best Local Similarity: 30.63% Mismatches: 155
Query Match: 12.26% Indels: 70
DB: 29 Gaps: 11
US-09-712-338-2_COPY_19_555 (1-537) x CNS06EAK (1-1005)
QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerThrValAspGlySerIle 145
Db 1 TTCTGGACACCCCTAGT-GTTGGTTTTCATATGGTGAT----- 41
QY 146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165
Db 41 ----- 41
QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaTrpGluIle 185
Db 42 -----GAAAGGTTACCAATACAGAGCTGCTGGTAGGACGCTCTACATCTTC 89
QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerIleAspPhe 205
Db 90 TTGGAATTATCTTCGAAAGGTTCCACATTG-----AGATCGCATGACTTC 137

QY 206 SerLeuThrThrGluSerTyrGlyHisTyrGlyProAlaPheAsnHisPheTyr 225
Db 138 CACATTCCTGGTGAATCCTATGCTGGACATTACATTCCTCAGATTGCCATGAAATCGTT 197
QY 226 GluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSer 245
Db 198 ATCGAAATCCT-----GGAAGAACITTTGACCTAACCTCT 233
QY 246 LeuGlyIleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrIleProGluPhe 265
Db 234 GTTCTTATCGCAATGATATACCGAATCCCTCCATCAAAATGATTACTATCAGCCAAATG 293
QY 266 AlaVal---AsnAsnThrTyrGlyIleLeuAlaValAsnGluThrValTyrAsnTyrMet 284
Db 294 GCITGGCGTGAAGGTGGATATCCTCAACTTTTAACGTATGAAGATGTTTCTCAGATGAG 353
QY 285 LysPheAlaAsnGlnMetProAsnGlyCysGlnAspLeuIleSerThrCysLysGlnThr 304
Db 354 AAAAAYACAATCGT-----TGCGTAGCTTGAACCGTATCTCTTACGGTACA 401
QY 305 AsnArgThrAlaLeuAlaAspTyrAlaLeuCysAlaGluAlaThrAsnMetCysArgAsp 324
Db 402 AAATCTAAC-----ATTGCATGTGTAGCTGCTACTGTCATTCIGTGAATCT 446
QY 325 AsnValGluGlyProTyrTyrAlaPheAlaGlyArgGlyValTyrAspIleArgHisPro 344
Db 447 GTTACTGTGGGCATATCCCAAGAAACATGGCTGTGAATGTCTATGATATCAGGGGTCT 506
QY 345 TyrAspAspProThrPro-----ProSerTyrTyrAsnLysPheLeu 358
Db 507 TGTGAGGATATGATAGTAGTGCACCTGCTACTTTGGTATGAATATGTTGATGATTACATG 566
QY 359 AlaLysAspSerValMetAspAlaIleGlyValAsnIle---AsnTyrThrGlnSerAsn 377
Db 567 AACCAAGATACACAGGACGCGTTGGGTCAGATGTACACAATTAACCGGTGTAAAC 626
QY 378 AsnAspValTyrTyrAlaPheGlnGlnThrGlyAspPheValTrpProAsnPheIleGlu 397
Db 627 GATCAAGTATCTTGGGATTTGCCCTAACTGGGGATGGTCCCAAGCT---TTCCACAA 683
QY 398 AspLeuGluGluIleLeuAlaLeuProValArgValSerLeuIleTyrGlyAspAlaAsp 417
Db 684 TATGIGACAGAACTGGTGGATTTAAACATCCACATCTGCTCTACCGCTGGTGATAAAGAT 743
QY 418 TyrIleCysAsnTrpPheGlyGlyGlnAlaValSerLeuAlaAlaAsnTyrSerGlnAla 437
Db 744 TACATTTGTAAGTGGTGGTAAATAAAGCTTGGTCTGATAAAATACATTGGAGATACGGT 803
QY 438 AlaGlnPheArgSerAlaGlyTyrThrProLeuLysVal-----AsnGly 452
Db 804 GAAAGTACGAACT-----CTACCTCTAAAGGCGCTGGAAGTCAAAAGTACTGCTGT 854
QY 453 ValGluTyrGlyGluThrArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAla 472
Db 855 GAAAAATTTGGTGAAGTAAAC-ATACGGCGCTTTGACCTTCTTAAGGATTTACGATGCT 913
QY 473 GlyHisGluValProTyrTyrGlnProIleAlaSerLeuGlnLeuPheAsnArgThrIle 492
Db 914 GGACATGGTTCCTTATGACCAGCTGAAGCGCTCTTGAATGGTGAACGACTGGATT 973
QY 493 PheGly 494
Db 974 ACAGGG 979
RESULT 10
AK032197
LOCUS
DEFINITION
Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:643041i23 product:protective protein for
beta-galactosidase, full insert sequence.
AK032197
ACCESSION
VERSION
AK032197.1 GI:26328022
KEYWORDS
HTC; CAP trapper.

[illegible]

QY 49 ValLysSerTyrSerGlyTyrValAspThrSerProGluSerHisThrPhePheThrPhe 68
 Db 169 TTCGGCAATACTCCGGTACTCTCAGACATCCAGACTCCCAAGCATTCCACACTACTGGTTT 228
 QY 69 PheGluAlaArgHisAsnProGluThrAlaProIleThrLeuTrpLeuAsnGlyGlyPro 88
 Db 229 GTGGAGTCGCAAGACGACCAAGACAGCCCGGTGGTCTTGGCTTAAGGGGGTCC 288
 QY 89 GlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSer 107
 Db 289 GSCGTGCAGCTCGCTCGATGGCTGCTACAGACAGCGCCCTTCTCATCCAGCAGAT 348
 QY 108 -----ThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeu 125
 Db 349 GGTCTACCCCTGGAG-----TACACCCCTATCTTGGACCTGATGCAAGCTGGTG 402
 QY 126 PheLeuSerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIle 145
 Db 403 TATATCAGTCCCGCAGCTGGGGTGGCTTCTCTACTCGAT----- 444
 QY 146 AsnProValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrPro 165
 Db 444 ----- 444
 QY 166 ThrIleAspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaIlePheGluIle 185
 Db 445 -----GACAGATGTCAGTGACCAATGACACAGAGTGGCGGAGAACAAATTATGAGCC 498
 QY 186 LeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPhe 205
 Db 499 CTTAAAGACTTCTTCGGCTCTTTCGG-----GAATACAGAGACAAAC 540
 QY 206 SerLeuIlePhe-----GluSerTyrGlyHisTyrGlyProAlaPheAsnHis 223
 Db 541 AAATCTTTCTCCGACAGAGAGAGTATGCTGGCATCTACATCCCGCAC----- 588
 QY 224 PheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPhe 243
 Db 589 -----TTGGCTGTACTGCTCAAGCAGGATCCTAGCATGAATCTT 627
 QY 244 AsnSerLeuGlyIleIleAsnGlyIleIle-----AspGluAlaIle 257
 Db 628 CAGGGCTGGCTGGGCAATGGACTTGCCTCCTATGACGAGACGACAACTCCCTGGTC 687
 QY 258 GluAlaProTyrTyrProGluPheAlaValAsnThrTyrGlyIleLeuAlaValAsn 277
 Db 688 TACTTTGCCCTACTACCATGGCTTCTGGGGAACAGACTTTGGACTTCACTGCAGACCCAC 747
 QY 278 GluThrValTyrAsnTyrMetLysPhe-----AlaAsnGlnMetProAsnGlyCysGluAsp 296
 Db 748 TGCTGGCTCAGACACAGTAGTAATCTTATGACAAACAGACCCAGAGTGTGAACAAT 807
 QY 297 LeuIleSerThrCysLysGlnThrAsnArgThrAlaLeuAlaAspTyrAlaLeuCysAla 316
 Db 808 CTCCTGGAAGTGTCTCGAATGTGGCAAAATCTGCCTCAACATCTACAACTCTATGCT 867
 QY 317 GluAlaThrAsnMetCysArgAspAsnValGluGlyProTyrTyrAlaPheAlaGlyArg 336
 Db 868 CCG-----TGTGCTGGTGGGTGGCCGCG----- 891
 QY 337 GlyValTyrAspIleArgHisProTyrAspAsp----- 347
 Db 892 -----ACACATAGATATGAGGACACACTGTAGTCCAGGATTTTGGCAAC 936
 QY 347 ----- 347
 Db 937 ATCTTCACTCGCTGCCACTTAAGCGGAGATTTCTTGGGACATGATCGCTTCTGGGAC 996
 QY 348 -----ProThrProProSerTyrTyr-----Asn 355
 Db 997 AAGGTACGCTTGGATCTCCTCGACCAACACAGACCGCCCTTCCAACTACCTCAACAAAC 1056

QY 356 LysPheLeuAlaLys-----AspSerVal-----MetAspAla 366
 Db 1057 CCTATGTTCCGAAGGCTCTCCACATCCCGAGTCCGCTGGCCGCTGGACATGTGCAAC 1116
 QY 367 IleGlyValAsnIleAsnTyrThr-----GlnSerAsnAspAspValTyrTyrAla 383
 Db 1117 TCTTGGTGAATTTACAGTACCGCGCTCTACCAAGAGCATGATCCAGTAC----- 1170
 QY 384 PheGlnGlnThrGlyAspPheValTrpProAsnPheIleGluAspLeuGluIleLeu 403
 Db 1171 -----CTGAAGCTGCTCAGI 1185
 QY 404 AlaLeuProValArgValSerLeuIleTyrGlyAspAlaAspTyrIleCysAsnTrpPhe 423
 Db 1186 TCACAGAAATCCAGATCTCTCTACACCGAGATGTGGACATGCGCTGCAACTCATG 1245
 QY 424 GlyGlyGlnAlaValSerLeuAlaAsnTyrSerGlnAlaAlaGlnPheArgSerAla 443
 Db 1246 GCGATGAGTGTGTGTGATTCGCTCAACACAGATGGAGTGCAGCGCGG----- 1299
 QY 444 GlyTyrThrProLeuLysValAsnGlyValGluTyrGlyGlu-----Thr 458
 Db 1300 -----CCGIGCTAGTGGACTACGGGAGAGCGGAGACAGTACGTGTTCTG 1350
 QY 459 ArgGluTyrGlyAsnPheSerPheThrArgValTyrGluAlaGlyHisGluValProTyr 478
 Db 1351 AAGGAGTGTTCACATCATCTCTCTCAACATCAAGGTGGGAGACACATGTCCTCCACG 1410
 QY 479 TyrGlnProIleAlaSerLeuGlnLeuPheAsnArg 490
 Db 1411 GACAAGCTCGAGCTGCTTTTACCAATGTTCTCGAGG 1446

RESULT 11

CNS06Z5E

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS06Z5E 963 bp DNA linear GSS 07-JUL-2001
 T3 end of clone AY0AA015G04 of library AY0AA from strain CBS 6340
 of Kluyveromyces thermotolerans, genomic survey sequence.

AL421896
 AL421896.1 GI:12205091
 GSS.

Kluyveromyces thermotolerans
 Kluyveromyces thermotolerans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

1 (bases 1 to 963)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bollen-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies

FEMS Lett. 487 (1), 3-12 (2000)
 20584711
 11152876

2 (bases 1 to 963)
 Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.
 and Dujon,B.

Genomic exploration of the hemiascomycetous yeasts: 10.
 Kluyveromyces thermotolerans

FEMS Lett. 487 (1), 61-65 (2000)
 20584720
 11152885

3 (bases 1 to 963)
 Genoscope.

Direct Submission

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

```

source
    1..963
        /organism="Kluyveromyces thermotolerans"
        /mol_type="genomic DNA"
        /strain="CBS 6340"
        /db_xref="taxon:4916"
        /clone="AY00A015G04"
        /clone_lib="AY00A"
        /note="end : T3"
misc_feature
    <270..>962
        /note="similar to Saccharomyces cerevisiae ORF YBR139w [
strong similarity to carboxypeptidase ]"
        /evidence=not_experimental
BASE COUNT      242 a   231 c   217 g   272 t   1 others
ORIGIN
Alignment Scores:
Pred. NO.:      2,42e-28      Length:      963
Score:          352.50      Matches:      88
Percent Similarity: 45.08%      Conservative: 31
Best Local Similarity: 33.33%      Mismatches: 98
Query Match:      12.21%      Indels:      47
Db:                29      Gaps:      5

US-09-712-338-2_COPY_19_555 (1-537) x CNS0625E (1-963)
Qy      1  LeuProGlySerThrProAlaSerValGly-----ArgArgGlnLeuProLysAsnPro 18
Db      297  GTCCCAAAATACCGTGCAGTCTATCAAAATGCGTGAATGCTGGAGAGTTCGGAGGTTGCTC 356
Qy      19  ThrGlyValLysThrLeuThrThrAlaAsnAsnValThrLeuArgTyrLysGluProGly 38
Db      357  ACGATAAGTGTGAGCTGTGATGACACCTACTCGATGAGGTAAGAGAGTGGACCCCTCG 416
Qy      39  AlaGluGlyValCysGluThrThrProGlyValLysSerTyrSerGlyTyrValAspThr 58
Db      417  AAGTGGGTGTC-----GATCTGTCAAGCAGTACTCGGGGTACTCGGAGTACTGGACATAC 464
Qy      59  SerProGluSerHisThrPhePheThrPhePheGluAlaArgHisAsnProGluThrAla 78
Db      465  GAGGATTCCAACACACTCTTTTACTGGGCATTTGAATCCAGAAACGACCCACTCAATGAC 524
Qy      79  ProIleThrLeuTrpLeuAsnGlyGlyProGlySerAspSerLeuIleGlyLeuPheGlu 98
Db      525  CCCGTTATTCTATGGCTGAACGGTGTCCAGGGTGTCTCTTTTCAAGGGGTATTATTTC 584
Qy      99  GluLeuGlyProCysHisValAsnSerThrPheAspAspTyrIleAsnProHisSerTrp 118
Db      585  GAATCTGGCCCTCTCCGTCGGCCCTGAGCTAAGCCCGTTCGCAATCCGATCCCTGG 644
Qy      119  AsnGluValSerAsnLeuLeuPheLeuSerGlnProLeuGlyValGlyPheSerTyrSer 138
Db      645  AATACAAATGCTACTGTCATATCTTGGAGCAGCCCTTAGCGGTGGTTTTCCTATGCT 704
Qy      139  AspThrValAspGlySerIleAsnProValThrGlyValGluValGluAsnSerSerPheAla 158
Db      705  GAC-----
Qy      159  GlyValGlnGlyArgTyrProThrIleAspAlaThrLeuIleAspThrThrAsnLeuAla 178
Db      708  -----GAGAGAGTGGCTCCACTCAACGACGAC 734
Qy      179  AlaGluAlaAlaTrpGluIleLeuGlnGlyPheLeuSerGlyLeuProSerLeuAspSer 198
Db      735  GGAAGGATGTTTCATATCTTGGAGTGTGTTTTCAGAGATGCCA-----782

```

```

Qy      199  ArgValGlnSerLysAspPheSerLeuTrpThrGluSerTyrGlyGlyHisTyrGlyPro 218
Db      783  CAGTTTAGATCTAAGCACTTCCATATTCAGTCAATCACTACGTCAGCATATCAATCCA 842
Qy      219  AlaPhePheAsnHisPheTyrGluGlnAsnGluArgIleAlaAsnGlySerValAsnGly 238
Db      843  -----GAGATTGCCCATCAATPCGCTGCTGCTCATGCAATCTGAC 881
Qy      239  ValGlnLeuAsnPheAsnSerLeuGlyIleIleAsnGlyIleIleAspGluAlaIleGln 258
Db      882  AAACCTTCATCTCACCCTCCATCATGATGTCACGGGATACAGATTCCTTAGTCCAG 941
Qy      259  AlaProTyrTyr 262
Db      942  TATGAMTACTAT 953

RESULT 12
AK019861
LOCUS
DEFINITION
    AK019861      2493 bp      mRNA      linear      HTC 05-DEC-2002
    Mus musculus 11 days pregnant adult female ovary and uterus cDNA,
    RIKEN full-length enriched library, clone:5031400F07
    product:protective protein for beta-galactosidase, full insert
    sequence.
ACCESSION
    AK019861
VERSION
    AK019861.1  GI:12860233
KEYWORDS
    HTC; CAP trapper.
SOURCE
    Mus musculus (house mouse)
ORGANISM
    Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1  Carninci,P. and Hayashizaki,Y.
        High-efficiency full-length cDNA cloning
        Meth. Enzymol. 303, 19-44 (1999)
        99279253
        PUBMED
        10349636
    2  Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
        Itoh,H., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
        Normalization and subtraction of cap-trapper-selected cDNAs to
        prepare full-length cDNA libraries for rapid discovery of new genes
        Genome Res. 10 (10), 1617-1630 (2000)
        20499374
        PUBMED
        11042159
    3  Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
        Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
        Sumi,N., Ishii,Y., Nakamura,S., Hazawa,M., Nishine,I., Harada,A.,
        Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
        Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,
        Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
        Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
        RIKEN integrated sequence analysis (RISA) system--384-format
        sequencing pipeline with 384 multicapillary sequencer
        Genome Res. 10 (11), 1757-1771 (2000)
        20530913
        PUBMED
        11076861
    4  Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
        Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
        Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
        Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R.,
        Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
        Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
        Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G.,
        Quackenbush,J., Schriml,L.M., Stauble,F., Suzuki,R., Tomita,M.,
        Wagner,I., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
        Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
        Carninci,P., de Bonaldo,M.P., Brownstein,M.D., Bult,C.,
        Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
        Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
        Marchionni,L., Mashima,J., Mazzarelli,J., Monbaerts,P., Nordone,P.,
        Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,

```



```

RESULT 15
CD045123
LOCUS
DEFINITION
    psHB017xN05f_183846 psHB: Infected hypocotyl soybean host. 48 hrs
    post infection Phytophthora sojae cDNA clone SHB017N05 5, mRNA
ACCESSION
CD045123
VERSION
CD045123.1 GI:30498716
KEYWORDS
EST.
SOURCE
Phytophthora sojae
ORGANISM
Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
REFERENCE
1 (bases 1 to 712)
  Tyler,B.M., Judelson,H.S., Glizzen,M., Dean,R.A. and Waugh,M.E.
  USDA-IFAFS: Expression of Phytophthora sojae genes during infection
  and propagation
JOURNAL
Unpublished
COMMENT
Contact: Tyler B
  Tyler lab
  VBI
  1880 Pratt Dr., Blacksburg, VA 24061, USA
  Tel: 540-231-7318
  Email: bmtyle@vt.edu
PCR Primers
  FORWARD: BK reverse
  Plate: 017 row: N column: 05
  Seq primer: BK reverse
  High quality sequence stop: 712.
FEATURES
    source
    1. 712
        /organism="Phytophthora sojae"
        /mol_type="mRNA"
        /db_xref="taxon:67593"
        /clone="SHB017N05"
        /tissue_type="infected host tissue"
        /cell_line="P6497"
        /dev_stage="48 hour post infection"
        /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
        post infection"
        /note="vector: pBK-CMV; site_1: EcoRI; site_2: XhoI;
        USDA-IFAFS: Expression of Phytophthora sojae genes during
        infection and propagation."
BASE COUNT
    154 a 224 c 218 g 116 t
ORIGIN
Alignment Scores:
Pred. No.: 2.7e-26 Length: 712
Score: 332.50 Matches: 82
Percent Similarity: 44.65% Conservative: 39
Best Local Similarity: 30.26% Mismatches: 97
Query Match: 11.52% Indels: 53
DB: 14 Gaps: 8

US-09-712-338-2_COPY_19_555 (1-537) x CD045123 (1-712)
QY 10 GlyArgArgGlnLeuProLysAsnProThrGlyValLysThrLeuThrThraAlaAsn 29
   |||
   |||
   |||
Db 42 GGCTGGCGCGCATCGAGACAGTGGCGGGTGGCTGCTACCGTCACTGCTGGCGGGGAC 101
   |||
   |||
   |||
QY 30 ValThrTleArgTyrLysGluProGlyAlaGluGlyValCysGluThrThrProGlyVal 49
   |||
   |||
   |||
Db 102 ACAATC-----ATGTGTGGCACC-----GCA 122
   |||
   |||
   |||
QY 50 LysSerTyrSerGlyTyrValAspThr-----SerProGluSerHisThrPhePheIrp 67
   |||
   |||
   |||
Db 123 ACCAACAGAGGGGCTACGTACCTCCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 182
   |||
   |||
   |||
QY 68 PhePheGluAlaArgHisAsnProGluThrAlaProIleThrLeuThrLeuAsnGlyGly 87
   |||
   |||
   |||
Db 183 TTCCTCGAGTCCCGCAGAGCTCTGCTACCGACCCCTTGGTGTCTGCTGCTGCTGCTGCTG 242
   |||
   |||
   |||

```

```

QY 88 ProGlySerAspSerLeuIleGlyLeuPheGluGluLeuGlyProCysHisValAsnSer 107
   |||
   |||
   |||
Db 243 CCCGGAGTTTCCAGCCTCATGACGCTCTTGACCGAAGACGCGCGTGTCTGTTAAGGAG 302
   |||
   |||
   |||
QY 108 ThrPheAspAspTyrIleAsnProHisSerTrpAsnGluValSerAsnLeuLeuPheLeu 127
   |||
   |||
   |||
Db 303 GATATGTGCGAGCGGAGCCCAACCTCGTGAACCTCGGAGGCAAGCTCATCTGGCTG 362
   |||
   |||
   |||
QY 128 SerGlnProLeuGlyValGlyPheSerTyrSerAspThrValAspGlySerIleAsnPro 147
   |||
   |||
   |||
Db 363 GACGACCCACGACGTCGGGTACGTCGACGCGAGCCCGCAGATGCGGAC---CACGAC 419
   |||
   |||
   |||
QY 148 ValThrGlyValValGluAsnSerSerPheAlaGlyValGlnGlyArgTyrProThrIle 167
   |||
   |||
   |||
Db 420 GAGAGGAGCGTCCAGGAGAAC----- 440
   |||
   |||
   |||
QY 168 AspAlaThrLeuIleAspThrThrAsnLeuAlaAlaGluAlaAlaThrPgluIleLeuGln 187
   |||
   |||
   |||
Db 441 -----GTCTACGCGTCTCTACAG 458
   |||
   |||
   |||
QY 188 GlyPheLeuSerGlyLeuProSerLeuAspSerArgValGlnSerLysAspPheSerLeu 207
   |||
   |||
   |||
Db 459 CGGTTTCGGACACACACCGGAGCTT-----CAAGCGGCGCGCTGTTCCTC 506
   |||
   |||
   |||
QY 208 TrpThrGluSerTyrGlyHisTyrGlyProAlaPhePheAsnHisPheTyrGluGln 227
   |||
   |||
   |||
Db 507 GCAGCGAGAGCTACGCGAGGCCACTACATCCAGCAGCGCGGCACAAAGATCCACCGGAG 566
   |||
   |||
   |||
QY 228 AsnGluArgIleAlaAsnGlySerValAsnGlyValGlnLeuAsnPheAsnSerLeuGly 247
   |||
   |||
   |||
Db 567 AACAAA-----GCGCGGAAGTCAAGGCGTCTAAACCTCCAGGCGCATCGCC 611
   |||
   |||
   |||
QY 248 IleIleAsnGlyIleIleAspGluAlaIleGlnAlaProTyrTyrProGluPheAlaVal 267
   |||
   |||
   |||
Db 612 ATCGGGAACGGCTTCACCAACACCGCTCGTCGAGCAGCGACACGGGCTGGACATG---GTC 668
   |||
   |||
   |||
QY 268 AsnAsnThrTyrGlyIleLysAlaValAsnGlu 278
   |||
   |||
   |||
Db 669 AACAACTCGTACGCGCTCAAGCTGATGGACGAC 701

```

Search completed: September 17, 2003, 01:44:29
Job time : 3392 secs